

RESULT	1	
Q9BU47		
ID	Q9BU47	PRELIMINARY; PRT; 321 AA.
AC	Q9BU47;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	HYPOTHETICAL 35.0 KDA PROTEIN.	
OS	Homo sapiens (Human) .	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=LUNG CARCINOMA;	
RA	Strausberg R.;	
RL	Submitted (FEB-2001) to the EMBL/genbank/DBJ databases.	
DR	EMBL; BC002894; AAH02894.1; -.	
DR	HSSP; P35555; IEMN.	

DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR SMART: SM00181; EGF; 3.
 DR SMART: SM00179; EGF_CA; 2.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00261; FU; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01187; EGF_2; 2.
 DR PROSITE: PS01187; EGF_CA; 2.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;
 KW Hypothetical protein; Repeat.
 SQ SEQUENCE 321 AA; 34958 MW; A74360A1D817F23D CRC64;

Query Match 89.1%; Score 1787; DB 4; Length 321;
 Best Local Similarity 90.9%; Pred. No. 9, 1e-164;
 Matches 321; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

OY 1 MRPPRAALGLPLLLPPAPAAKPPCHRCGLVDKFNQGMVDTAKNFGGNTAW 60
 DB 1 MRPPRAALGLPLLLPPAPAAKPPCHRCGLVDKFNQGMVDTAKNFGGNTAW 60
 OY 61 EEKTLKSESEIRLEILGLCESSDFECNMLEAOEHLAMWLQLKSEYDLPFEMFC 120
 DB 61 EEKTLKSESEIRLEILGLCESSDFECNMLEAOEHLAMWLQLKSEYDLPFEMFC 120
 OY 121 VKTLKCCSPGTGPPDLACQGGSORPCSGNGHCSGDSHQDGSCHNGYGGPLCTDC 180
 DB 121 VKTLKCCSPGTGPPDLACQGGSORPCSGNGHCSGDSHQDGSCHNGYGGPLCTDC 180
 OY 181 MDGYFSLRNETHSICITACDESKCTCSGLTNRCCGCEGVWVLDGACVYDCAEAPP 240
 DB 181 MDGYFSLRNETHSICITACDESKCTCSGLTNRCCGCEGVWVLDGACVYDCAEAPP 240
 OY 241 CSAAOFCKNANGSYTCEDCSGCVGTGEGPGNCKECISGYAREHGCADVDDECSLAEKT 300
 DB 241 CSAAOFCKNANGSYTCEDCSGCVGTGEGPGNCKECISGYAREHGCADVDDECSLAEKT 300
 OY 301 CVRKNENCVTPGSIYCVCPDGEETEDACVPAAEAATGESPTLPSPREDL 353
 DB 269 CVRKNENCVTPGSIYCVCPDGEETEDACVPAAEAATGESPTLPSPREDL 321

RESULT 2
 OQCYAO PRELIMINARY; PRT; 350 AA.
 AC OQCYAO;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 5730592L2LRK PROTEIN.
 GN 5730592L2LRK.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schraml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazarrelli J., Mombeerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK017880; BAB30986.1; -.
 DR HSSP; P35555; 1EMN.
 DR MGD; MGI:1923987; 5730592L2LRK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR002174; Furin-like.
 DR SMART: SM00181; EGF; 4.
 DR SMART: SM00179; EGF_CA; 3.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00261; FU; 3.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 KW EGF-like domain; Glycoprotein; Hydroxylation.
 SQ SEQUENCE 350 AA; 38219 MW; 781D7389B1944231 CRC64;

Query Match 76.5%; Score 1533.5; DB 11; Length 350;
 Best Local Similarity 75.6%; Pred. No. 2, 4e-139;
 Matches 267; Conservative 30; Mismatches 53; Indels 3; Gaps 2;

OY 1 MRPPRAALGLPLLLPPAPAAKPPCHRCGLVDKFNQGMVDTAKNFGGNTAW 60
 DB 1 MHLLAARGL--LILPPGAVASRRPTMCQCRFLVYKFNQGMVDTAKNFGGNTAW 58
 OY 61 EEKTLKSESEIRLEILGLCESSDFECNMLEAOEHLAMWLQLKSEYDLPFEMFC 120
 DB 59 EEKTLKSESEIRLEILGLCESSDFECNMLEAOEHLAMWLQLKSEYDLPFEMFC 118
 OY 121 VKTLKCCSPGTGPPDLACQGGSORPCSGNGHCSGDSHQDGSCHNGYGGPLCTDC 180
 DB 119 VKTLKCCSPGTGPPDLACQGGSORPCSGNGHCSGDSHQDGSCHNGYGGPLCTDC 178
 OY 181 MDGYFSLRNETHSICITACDESKCTCSGLTNRCCGCEGVWVLDGACVYDCAEAPP 240
 DB 179 TDGFSLRNETHSICITACDESKCTCSGPNKCIQCEVGMVAVEDACVYDCAEAPP 238
 OY 241 CSAAOFCKNANGSYTCEDCSGCVGTGEGPGNCKECISGYAREHGCADVDDECSLAEKT 300
 DB 239 CSDFGVCEVNGSYTCEDCSGCVGTGEGPGNCKECISGYAREHGCADVDDECSLAEKT 298
 OY 301 CVRKNENCVTPGSIYCVCPDGEETEDACVPAAEAATGESPTLPSPREDL 353
 DB 299 CVRKNENCVTPGSIYCVCPDGEETEDACVPAAEAATGESPTLPSPREDL 350

RESULT 3
 OQ0438 PRELIMINARY; PRT; 348 AA.
 AC OQ0438;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HT PROTEIN.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxId=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chen H., Okubo, T., Ling V., Zhang W.;

Submitted (Feb-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U048852; AAA91469.1; .
 DR HSP: P35555; 1EM.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002049; Laminin-EGF.
 DR InterPro: IPR001005; Myb_DNA_bind.
 DR SMART: SM00179; EGF_CA_1.
 DR SMART: SM00001; EGF_Like_2.
 DR SMART: SM00261; FU; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01187; EGF_CA; 2.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SO SEQUENCE 348 AA; 38199 MW; BD61F6C89971B86D CRC64;

Query Match 73.9%; Score 1481.5; DB 11; Length 348;
 Best Local Similarity 73.9%; Pred. No. 2,4e-134;
 Matches 255; Conservative 31; Mismatches 58; Indels 1; Gaps 1;

OY 1 MRLPRAALGLPLLLPAPPAEAKKPTPCRCGLVDKFGNOGMDATKKNFGGNTAW 60
 DB 1 MHLPPAAVGLL-LLLPAPPAVASKPTMCQRCRALVDKFNQGMANTARKNFGGNTAW 59
 OY 61 EEKLTSSYESSERLELEIGLCESSDFECNOMLEAOEHLAWMLQKSEYPLDFEWC 120
 DB 60 EEKLSKYERSEIRLEIMEGLCDSNDFECNQLEQHEQLEAWMQLTKKRCPLFEWC 119
 OY 121 VKTKLVCCSGTYPPDLCAQGGSGRSGNGHSGSGSGSGSGSGSGSGSGSGSGSG 180
 DB 120 VHTLKACCLGTGTGPDQECQGGSGRSGNGHSGSGSGSGSGSGSGSGSGSGSGSG 179
 OY 181 MDGFSSLRNETHSICACDESKTSGLTNRDCGECEVGVNLDEGACVDVDECAAEPP 240
 DB 180 MDGFSSLRNETHSICACDESKTSGLTNRDCGECEVGVNLDEGACVDVDECAAEPP 239
 OY 241 CSAAGFCNNAGSYTCCECDSSCVGTGEGPNCCEKISGVARHGGACADVDECSLA 300
 DB 240 CSNAYQCEENAGSYTCCECDSTGCTGKGPANCKECISGYSKQKGCADIDCSLETKV 299
 OY 301 CVRNENCTYTPGSYVCVCPDGEETEDACVPAEAETGESEFT 345
 DB 300 CKKENENCTYTPGSYVCVCPDGEETEDACVPAEAETGESEFT 344

RESULT 4
 O96HDI PRELIMINARY; PRT; 422 AA.
 AC O96HDI;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE UNKNOWN (PROTEIN FOR MGC:8447).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG CARCINOMA;
 RA Strausberg R.;
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC008720; AAH08720.1; .
 SO SEQUENCE 422 AA; 45909 MW; C26CB6EF7052E1B5 CRC64;

Query Match 48.3%; Score 969; DB 4; Length 422;

Best Local Similarity 47.1%; Pred. No. 5.8e-85;
 Matches 168; Conservative 50; Mismatches 119; Indels 20; Gaps 4;

OY 7 AALDLPLULLLP-----PAPBA--KKPTCHRCRGVLDKFNQGMVDTAKKNFGG 56
 DB 12 AVMGLSLFLNLPPIWLPSPPOSSPPQPHCHCRGLVDKFNKLEFTINDNFGG 71
 OY 57 NTAMEEKLTSKYESSERLELEIGLCESSDFECNOMLEAOEHLAWMLQKSEYPLDF 116
 DB 72 NTAMEEKLTSKYESSERLELVLEGVCSKSPFCHRLLESEELVESMWFKQDEADPLF 131
 OY 117 EMFCVKTLKYCSBPGTYGPDCLACQGSQRPSCNGHSGSGSGSGSGSGSGSGSGSG 176
 DB 132 QWLSDSLKLKCPAGTGTGSPCLPCPGTERPCGGVGGQCEGRTGSGSCHDCQAGYGEA 191
 OY 177 CTDCMDGYFSSLRNETHSICACDESKTSGLTNRDCGECEVGVNLDEGACVDVDECA 236
 DB 192 CGQGLGYFEAERNASHLVCSACFGPCARCSGPESNLOCKGMALNHLKCVIDECGT 251
 OY 237 EPPCSAAGFCNNAGSYTCCECDSSCVGTGEGPNCCEKISGVARHGGACADVDECSL 296
 DB 252 EGANGCADQFCVNTGSETECHDCAKACLGAGAPGRCKKCSPTGYQVGSCLVDDECE 310
 OY 297 AEKTCVRKNENCTYTPGSYVCVCPDGEETEDACVPAEAETGESEFTPLPSREDL 353
 DB 311 -TEVCGPENKCCENTEGEGRCTICAGYKQMEGICV-----KEQIPGAPPIITDL 358

RESULT 5
 O91XD7 PRELIMINARY; PRT; 420 AA.
 AC O91XD7;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE UNKNOWN (PROTEIN FOR MGC:18896).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC010804; AAH10804.1; .
 SO SEQUENCE 420 AA; 45717 MW; 406BF2D739D3179 CRC64;

Query Match 48.3%; Score 968.5; DB 11; Length 420;
 Best Local Similarity 47.0%; Pred. No. 6.4e-85;
 Matches 171; Conservative 43; Mismatches 121; Indels 29; Gaps 5;

OY 3 LPRRAALGLPLULLL-----PAPBAAKKPTPCRCRGVLDKFNQGM 45
 DB 4 LPPR---GLVPSLLMCLSLFLSPGVWLPSPRPHPRAERPHPCRTALVDNFNKGL 60
 OY 46 VDTAKKNFGGNTAMEEKLTSKYESSERLELEIGLCESSDFECNOMLEAOEHLAW 105
 DB 61 ERTIRDNFGGNTAMEEKLTSKYDSEIRLVLEGVCSRSDFCHRLLESEELVENMW 120
 OY 106 LQLSEYVPLDFEMFCVKTLKYCSBPGTYGPDCLACQGSQRPSCNGHSGSGSGSGSGSG 165
 DB 121 FHRQEARPLDFQWLCSDBSLKLCPSGTGTPCLPCPGTERPCGGYGGCBEGTRGSGH 180
 OY 166 CRCHMGYOGPLCTDCMDGYFSSLRNETHSICACDESKTSGLTNRDCGECEVGVNLDE 225
 DB 181 CDCQAGYGEAGCGGLGYFEAERNASHLVCSACFGPCARCTGPEESHCLQCKKGMALHN 240
 OY 226 GACVDVDECAAEPPCSAAGFCNNAGSYTCCECDSSCVGTGEGPNCCEKISGVARH 285
 DB 241 LKCVIDECGTGATCGADQFCVNTGSETECHDCAKACLGAGAPGRCKKCSPTGYQV 300

QY	286	GCCAGVDECSLAEKFCVKNKNCNYPGSGVVCVCPDGFEEETDAC----	VPAAE----	AA	338
DB	301	SKCLDVDECEY--VCPGENKCENTFEGGRCVCAEGIKQEDGICVKEOVPESAGFAEM			358
QY	339	TEGE 342			
DB	359	TEDE 362			
RESULT	6				
Q9Y409	09Y409	PRELIMINARY;	PRT;	417	AA.
AC	Q9Y409:				
DT	01-NOV-1999	(TREMBLrel. 12, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	HYPOTHETICAL. 44.9 KDA PROTEIN.				
GN	DKF2P566D213.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=KIDNEY.				
RX	MEDLINE=21154917; PubMed=11230166;				
RA	Wleminan S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,				
RA	Ansorge W., Boecker M., Bloecker H., Bauersachs S., Blum N.,				
RA	Labber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,				
RA	Mewes H.W., Oltmawelder B., Obermair B., Tampe J., Heubner D.,				
RA	Wambutt R., Korn B., Klein M., Poustka A.;				
RT	"Towards a Catalog of Human Genes and Proteins: Sequencing and				
RT	Analysis of 500 Novel Complete Protein Coding Human cDNAs.;"				
RL	Genome Res. 11:422-435(2001).				
DR	EMBL; AL050275; CAB3376.1; -.				
DR	HSSP; P35555; 1EKN.				
DR	InterPro; IPR000152; Asx_hydroxyl.				
DR	InterPro; IPR000561; EGF-like.				
DR	InterPro; IPR001881; EGF_Ca.				
DR	InterPro; IPR002174; Furin-like.				
DR	InterPro; IPR002049; Laminin_EGF.				
DR	SMART; SM000181; EGF_1.				
DR	SMART; SM00001; EGF-like; 1.				
DR	SMART; SM00261; FU; 2.				
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.				
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.				
DR	PROSITE; PS01186; EGF_2; 1.				
DR	PROSITE; PS01187; EGF_Ca; 2.				
DR	PROSITE; PS01248; LAMININ_Type_EGF; UNKNOWN_1.				
KW	Calcium-binding; EGF-like domain; Glycoprotein; Hypothetical protein;				
KW	Repeat.				
SO	SEQUENCE	417	AA;	44934	MM; C2E38C3714E5C4B8 CRC64;
Query Match		47.9%;	Score 959.5;	DB 4;	Length 417;
Best Local Similarity		47.9%;	Pred. No. 4.7e-84;		
Matches 169;	Conservative	47;	Mismatches 116;	Indels 19;	Gaps
QY	7	AALGLPLLLPL-----PAPAA--KRP	TPCHRGVLDKFNQGVWDPAKKF	GGG	56
DB	12	AVLMKLSLEFLNPGPIWLOPPPOSSPPPOPHCHTCGVLDSFNGKLGERTIRDFGGG	71		
QY	57	NTAWBEKTLKSYSEIRILLLELLEGLCESSDFECNOMLEPAOEHLAEVWVLDKSEYPLD	116		
DB	72	NTAWBEENLSKKYKDESTRVLEVLVEGVCSKSDPCHRLLELSEELVESWFFHQGAPDLF	131		
QY	117	EMFCYKTLKVCSPGTGPDCLACOGGSGORPCSGNGCHSGGDSRGDGSRCRHMGTQGL	176		
DB	132	QWLCSDSLTCLCPACTGFGSCLPFCGGTERPFGGVCQEGEGERTFGSGCHDCQAGYGBA	191		
QY	177	CTDMDGVESSLRNFTHSICTACDESSCKKCSGLTIRNDGCECEVAVWLDGACVNDVECA	236		
DB	192	CGQGLGTFPEARARNSHLVCSACFGPCARCSPBESNCTLQCKGVALHLHKVLDDECGT	251		

QY	237	EPFCSAOCNKNANGNYTCGEEEDSCVCGTCGGPCNCECTJSGAREHGCAADDECSL	296
Db	252	EGANCGADGCPVCTVNGESTYEFCDRAKACLGMGAGPFRCKKCSPG10QVSKCLUDBE-	310
QY	297	AERTCVKRNKNCYNTPGSYVCVCPDGFETEDACVP---PAEA---EATEGE	342
Db	311	-TEVCPEGNOCENTGEGYRCICAEQCKQMEGICVMEQJPSAGFSEWTEDE	362
RESULT	7		
Q9VJP0	Q9VJP0	PRELIMINARY;	PRT; 374 AA.
AC	Q9VJP0;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	CG11377 PROTEIN.		
GN	CG11377.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY.		
RX	MEDLINE=20196006; PubMed-10731132.		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	Georg J.C., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Anderson M.D., Zhang L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,		
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abail J.F., Agbayani A., An H.-J., Andrews-Ffiankoch C., Baldwin D.,		
RA	Ballew R.M., Basu P.V., Bertram B.P., Bhandari D., Bolshakov S.,		
RA	Beeson K.Y., Benos P.V., Bouck J., Brokstein P., Brotlier P.,		
RA	Borkova D., Botchan M.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Burtis K.C., Busam D.A., Butler K., Davenport L.B., Davies P.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Duan A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Mayas-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Hout J.,		
RA	Hoslin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,		
RA	Jatani B., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paebel J.M.,		
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Strizbas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,		
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang S., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhu Q., Zheng L.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;		
RT	"The genome sequence of Drosophila melanogaster.";		
RL	Science 287:2185-2195(2000).		
DR	EMBL; AE003590; AAF51560.1; -		
DR	FLYBase; FBgn0031217; CG11377.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR001881; EGF Ca.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR002049; LamAhn_EGF.		
DR	SMART; SM00181; EGF; 1.		

DR SMART; SM00001; EGF_1like; 1.
DR SMART; SM00261; FU; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 374 AA; 41161 MW; 6B093AFC28331B0D CRC64;

Query Match 33.4%; Score 670; DB 5; Length 374;
Best Local Similarity 40.5%; Pred. No. 2,8e-56;
Matches 122; Conservative 48; Mismatches 109; Indels 22; Gaps 8;

QY 28 PTCCHRCGLVDFKFNQGMVDTAKKNGGNTAMEEKLSTKSEETILLETLECLSSD 87
DB 28 PTPCRACCTQLVSSFRAL-ERTKRGHAGDPTAMEEKLSTKSEVRLVLEQELCEGE 86
QY 88 F---ECNOMLEAOEHLLEAMWLQKSEYDPLEFEMFCVTKLVKCCSPGTYGPDCLACGG 143
DB 87 VINKDHCHLANEHEALLEDMFTIKQTESPDLSGLCTIDQLTYCCPLNTYGPDLCTC--- 143
QY 144 SQPCSGNGHCGSGDGSROGDSRCRHMGYOGPLCTDCMDGYFSSLRNETHSICTACDESC 203
DB 144 --TECNNGKCKGDKTRKNGKCKCDPGYAGPNCNECGPEHYEFSDKEKLLCTQCHAC 201
QY 204 KT--CSGLTRNDGCEGVAVLD-EGACVDVDECAE--PPPCAAPCPCKNANSTYCEE 258
DB 202 GEGGCTGEGGKSKCKCKGKMSMSEACVINECLEEQRPCKPQPCVNNESFSFCL 261
QY 259 CDSGVCCTEGPGNCKECSISYAREHGQCADVDECSLAERTCYRKNENCTNPGSYVCV 318
DB 262 CDRSCDCGDDGPRMCKKCKADGYELKRGKCHDIS---AEG---RSNVSTRTLLTYFGM 314
QY 319 C 319
DB 315 C 315

RESULT 8
Q9DFE9 PRELIMINARY; PRT; 165 AA.
AC Q9DFE9;
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 17, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOHETICAL 18.7 KDA PROTEIN (FRAGMENT).
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RA Bayne C.J., Gerwick L., Fujiki K., Nakao M., Yano T.;
RT "Immune-relevant (including acute phase) genes identified in the
RT livers of rainbow trout, Oncorhynchus mykiss, by means of suppression
RT subtractive hybridization."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF81341; AAG30016.1; -
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00181; EGF_1.
DR SMART; SM00001; EGF_1like; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER 1 165
FT TER 165
SQ SEQUENCE 165 AA; 18682 MW; A93E87C38CF4502 CRC64;

Query Match 29.9%; Score 599.5; DB 13; Length 165;
Best Local Similarity 57.0%; Pred. No. 6.7e-50;
Matches 94; Conservative 28; Mismatches 42; Indels 1; Gaps 1;

QY 82 LCSSDFECNOMLEAOEHLLEAMWLQKSEYDPLEFEMFCVTKLVKCCSPGTYGPDCLACQ 141
DB 1 LCSSDFECNOMLEAOEHLLEAMWLQKSEYDPLEFEMFCVTKLVKCCSPGTYGPDCLACQ 60
QY 142 GGSORPCSGNGHCGSGDGSROGDSRCRHMGYOGPLCTDCMDGYFSSLRNETHSICTACDE 201
DB 61 GGSERPCGHNGVCGDGTGNGNRCNDHGYKGEFCLDCMDGYFSSLRNETHSICTACDE 120
QY 202 SCTKSGLTNRDCEGVAVL-VIDEGACVDVDECAEPPPCAQAQ 245
DB 121 SCTKSGLTNRDCEGVAVL-VIDEGACVDVDECAEPPPCAQAQ 165

RESULT 9
Q19267 PRELIMINARY; PRT; 356 AA.
AC Q19267;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F09E8.2 PROTEIN.
GN F09E8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99065613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z73896; CAA98055.1; -
DR InterPro; IPR00152; ASX_Hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000004; Sapp.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_1like; 1.
DR PROSITE; PS00010; ASX_Hydroxyl; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 356 AA; 39790 MW; EFF82577DE334F57 CRC64;

Query Match 26.9%; Score 540; DB 5; Length 356;
Best Local Similarity 36.1%; Pred. No. 8.5e-44;
Matches 108; Conservative 43; Mismatches 110; Indels 38; Gaps 6;
QY 11 LPLPLLLPAPPAKPTPCRHRCGLVDFKFNQGMVDTAKKNGGNTAMEEKLSTKSEY 70
DB 7 LPLPLLLPAPPAKPTPCRHRCGLVDFKFNQGMVDTAKKNGGNTAMEEKLSTKSEY 66
QY 71 SEIRLLETLEGLCSS-----DFECNOMLEAOEHLLEAMWLQKSEYDPLE 116
DB 67 SETRLIEVLEGVCKKSSLPNNDNFMGIAIEFKSCTOLEKHEETIEFY--YNQOHNM 124
QY 117 EMPCVTKLVKCCSPGTYGPDCLACGGSORP--CSGNGHCGSGDGSROGDSRCRHMGYOG 174
DB 125 NWLCEVQLKLCPPGHRKNEQCPGLSEKADVCFGKSGCHGDSRSGKCKETGTG 184
QY 175 PLCTDCMDGYFSSLRNETHSICTACDESC-KTCSGLTRNDGCEGVAVLDEGACVDVDE 233

```

Db 185 NLCRYDIEFEESRTVOGVYCKKHGBCLGVSSSESSKCKCKNGWKLTEBGCADVNE 244
OY 234 CAAEPPCSAACPCKNANGSYTCCECDSSCGCTGEGPNCCKICISGAYAREHQCADVD 292
Db 245 CONESACTKHELCVMTVGSFKCE-----CKE---GYKKDDQONCOFD 284

RESULT 10
ID 073774 PRELIMINARY; PRT; 704 AA.
AC 073774;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FIBULIN-1, ISOFORM D.
GN FBLN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=99120531; PubMed=9923656;
RA Barth J.L., Agraves K.M., Roark E.F., Little C.D., Agraves W.S.;
RT "Identification of chicken and C. elegans fibulin-1 homologs and
RT Matrix Biol. 17:635-646(1998).";
RL EMBL: AF051399; AAC05387.1; -.
DR HSSP: P00742; IHG.
DR InterPro: IPR000020; Anaphylatoxin.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF01821; ANATO; 2.
DR Pfam: PF00008; EGF_6.
DR SMART: SM00104; ANATO; 3.
DR SMART: SM00179; EGF_CA; 8.
DR SMART: SM00001; EGF_Like; 1.
DR PROSITE: PS01177; ANAPHYLATOXIN_1; UNKNOWN_1.
DR PROSITE: PS01178; ANAPHYLATOXIN_2; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 5.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01187; EGF_CA; 8.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 704 AA; 78137 MW; D47D5A30D5E42932 CRC64;

Query Match 14.3%; Score 286.5; DB 13; Length 704;
Best Local Similarity 25.7%; Pred. No. 4,4e-19;
Matches 127; Conservative 37; Mismatches 120; Indels 211; Gaps 29;

OY 6 RAALGLPLLLPLPPAPEAKKPTPCRCRGVLKFNQGVDPAPKKNFGGNTAMEKTL 65
Db 5 RGRAPRLLLLLALLPALRGDLSMEEC---CDK---GVENANKN----- 44

OY 66 SKYESSEIRLLELEGLICSSDFECNQMLEAQ-----EEHLEAMWLQIKSEYPDL---- 115
Db 45 -----RICTSLPLISER--ECS-MTQVQCCKSLKEHYHSCDIEFASVHEEDSHN 93

OY 116 -----FMF-----CVYTLKV-----CSPGYIGPDL- 138
Db 94 GENSTCEAFYFKKCYCCLLGKTAQVQGSCEFNLRIGYQGITVAFRCCKVGGEGTDVSI 153

OY 139 -----ACOGGSRPCSGNGHSGDGRSGDSCRCRHMGY 172
Db 154 SDDAPKKEQVEISKELDQEDPILHDCRKG--PCS--QQCNDTSSY---VCSFVGI 206

OY 173 Q-----GPLCTD---CMDGYFSSLRNETHS--ICTAC-----DESCKCSGLTN- 211
Db 207 QLPDPVNCEDINECITG-----THSGIGQTCVNTLGSFRQGRDRDSCGIGYELTLD 258

OY 212 ---RDCGCEV-----GWLIDE--GACVDVDECAEPP 239

```

```

Db 259 SRCKDIDECETGTHNCPPDFICONTGSPRCRKLQCMNGFIODALGNCIDINECLSTNN 318
OY 240 PCSAOPCKNANGSYTC-----EECDSSCVCTGE-----G 270
Db 319 PCPAGQICINTDGSYTCORISPCGGRGYHLNEDGTACVDVBESSDDPC-GEHVCINC 377

OY 271 PGNCK-ECISGVARE--HGOCADVDEC-SLAERTCYRKNEHCYNTPGSYVVCPPDGEET 326
Db 378 PGNYRCECKSGSYFVDYISRCIDINECRYPGRCAHK---CENTPGSYVCTCTMFKLS 434

OY 327 ED--ACVPAEAT 339
Db 435 SDRSCEDINECESS 449

RESULT 11
ID 09BLJ1 PRELIMINARY; PRT; 937 AA.
AC 09BLJ1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CI-METAL.
GN CI-METAL.
OS Clona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OC Clonidae; Clona.
OX NCBI_TaxID=7719;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=21347414; PubMed=11455433;
RA Nakayama A., Satou Y., Satou N.;
RT "Isolation and characterization of genes that are expressed during
RT Clona intestinalis metamorphosis.";
RL Dev. Genes Evol. 211:184-189(2001).
DR EMBL: AB041857; BAB40596.1; -.
DR HSSP: P00742; IHG.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001212; Somatomedin_B.
DR InterPro: IPR001491; Thrombomodulin.
DR Pfam: PF00008; EGF_5.
DR PRINTS: PR00907; THROMBOMODULN.
DR SMART: SM00181; EGF_19.
DR SMART: SM00179; EGF_CA; 18.
DR PROSITE: PS00010; ASX_HYDROXYL; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 19.
DR PROSITE: PS01187; EGF_CA; 12.
DR PROSITE: PS00524; SOMATOMEDIN_B; UNKNOWN_1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 937 AA; 101043 MW; BC67B30CB8391D07 CRC64;

Query Match 13.9%; Score 279.5; DB 5; Length 937;
Best Local Similarity 29.3%; Pred. No. 2,9e-18;
Matches 80; Conservative 24; Mismatches 96; Indels 73; Gaps 14;

OY 121 VTKLVCC-SPTGY-----GPDLACQGSQRPCSGNGHSGDGRSGDSSCRC 168
Db 461 VNTNKRCTNTPGSRFCVNGRYRAQSRVVDINECRSSPCGNNAKCI---NTPGSETCRC 517

OY 169 HMGY--QGELCTDCMDGYFSSLRNETHSICTACDECKCSGLTNDGCECEYGVWLIDG 226
Db 518 NTGYTGNGLICRDI-----NECEAHNPNGENALCINGISYRC-ICARGF--SGP 565

OY 227 ACVDVDECAEPPCSAOPCKNANGSYTCCECDSSCGCTGEG- 270
Db 566 LCTDFNECAAIRPPCSPNADCTNTPGSEFC-QCKP---GYTGGLVCRDINECSRNPACP 621

```

QY 271 -----PG--NCKECISGYAREHGOCADVDECSLAETKCYRRKNENCYTPGSGVCVP 320
DB 622 RNCRCINTPGGFNC-VCAIGRYKRVGNOCVYDINECRASRRPC-DLNASQNTPGSFTCTCN 679
QY 321 DGEF-----ETEDACVPAAEATEG 341
DB 680 TGYTGNGLTCADINECNPRACHPOATCANTPG 712

RESULT 12
000508
ID 000508 PRELIMINARY; PRT; 1587 AA.
AC 000508;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LATENT TGF-BETA BINDING PROTEIN-4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97415399; PubMed=9271198;
RA Giltay R., Kostka G., Timpl R.;
RT "Sequence and expression of a novel member (LTBP-4) of the family of
RT latent transforming growth factor-beta binding proteins.";
RL FEBS Lett. 411:164-168(1997).
DR EMBL: Y13622; CAJ73944.1; -;
DR HSSP: P35555; IEMN.
DR Interpro: IPR000152; Asx_hydroxyl.
DR Interpro: IPR000561; EGF-like.
DR Interpro: IPR001881; EGF_Ca.
DR Interpro: IPR003018; GAF.
DR Interpro: IPR002212; TB.
DR Pfam: PF00008; EGF; 17.
DR Pfam: PF00683; TB; 4.
DR SMART: SM00179; EGF_CA; 15.
DR SMART: SM00001; EGF_Like; 5.
DR SMART: SM00065; GAF; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 12.
DR PROSITE: PS01187; EGF_CA; 17.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 1587 AA; 169620 MW; 57A832F95FA0AE46 CRC64;

Query Match 13.6%; Score 272; DB 4; Length 1587;
Best Local Similarity 24.0%; Pred. No. 2.8e-17;
Matches 113; Conservative 27; Mismatches 122; Indels 208; Gaps 27;

QY 19 PPAP-----BAKKRTPC--HRCRGLVDKEN-----QGNVDTAKNFGGNTAMEKTLISK 67
DB 709 PGARCDQVDECARSPRTYGRCENTEGSFQCVPMGFQF-----NTAGSE----- 754
QY 68 YESEIHLLEIGLCSSDFECNQMLAEQENLEAMWLQKSEYRDLFEWFCVTLKVC 127
DB 755 -----CEDVD-EC-----ENHLACPGQE-----CYN----- 774
QY 128 CSPGTYPDLACGG-----SORPGSGNGHCSGDGSRQDGSRCRHMG 171
DB 775 -SPGSF--QGTCPSGNHLHRCRTDVDECSGARPCGPHNCT--NTESFRCSCAPG 828
QY 172 YQGL-----CTD--CMDGYF-----SLRNETHSICTA-----CD 200
DB 829 YRASGRGRCADVNECLSEGFRCFHGECINTDGSFACTCAGYRPGRGASCLDVDECS 888
QY 201 ESKCTCSGL-TNRD-----CG----- 215
DB 889 EEDLCGSGICTNTDGSFECICPPGHRAGPDLASCLDVDECSERGRALCGSGRCENSPGSY 948
QY 216 ---ECEVGW-VLDEGACVDVDECAERPPCSAAQFCNANGSYTC--EEDCSGCGTGE 269

DB 949 RCVBDCDPRGYAGREGTDDVDECOXEYGEICGAQRCEMTGSGYRCFACDPGYQPTPG- 1007
QY 270 GPCKCKE-----CISGT--AREHGOCADVDECSLAETKC 301
DB 1008 --GGCQDVDECRNRSFCGAHNAVCONLPGSFQCLDQGYEGARDGHNHCYDVNECEFTLQVC 1065
QY 302 VRKNENCYTPGSGVCVCPDQGEETE--DACVPAEATEGSPQDLP 348
DB 1066 --GAALCENVEGSFLCYCPSNPEEPDPTGRCVPP--RTSVGMSPGSQP 1110

RESULT 13
075412
ID 075412 PRELIMINARY; PRT; 1511 AA.
AC 075412;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 4S.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=HEART;
RX MEDLINE=98325059; PubMed=9660815;
RA Saharinen J., Taipale J., Monni O., Keski-Oja J.;
RT "Identification and characterization of a new latent transforming
RT growth factor--binding protein, LTBP-4.";
RL J. Biol. Chem. 273:18459-18469(1998).
DR EMBL: AF051344; AAC39879.1; -;
DR HSSP: P35555; IEMN.
DR Interpro: IPR000152; Asx_hydroxyl.
DR Interpro: IPR000561; EGF-like.
DR Interpro: IPR001881; EGF_Ca.
DR Interpro: IPR002212; TB.
DR Pfam: PF00008; EGF; 17.
DR Pfam: PF00683; TB; 4.
DR SMART: SM00179; EGF_CA; 15.
DR SMART: SM00001; EGF_Like; 5.
DR PROSITE: PS00010; ASX_HYDROXYL; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 12.
DR PROSITE: PS01187; EGF_CA; 17.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 1511 AA; 161157 MW; C61AB7578256958D CRC64;

Query Match 13.5%; Score 271.5; DB 4; Length 1511;
Best Local Similarity 23.6%; Pred. No. 2.9e-17;
Matches 110; Conservative 29; Mismatches 131; Indels 197; Gaps 24;

QY 19 PPAP-----BAKKRTPC--HRCRGLVDKFNQGNVDTAKNFGGNTAMEKTLISKYESS 71
DB 633 PGARCDQVDECARSPRTYGRCENTEGSFQCVPMGFQPAAGS----- 678
QY 72 EIRLEILEIGLCSSDFECNQMLAEQENLEAMWLQKSEYRDLFEWFCVTLKVCSPG 131
DB 679 -----CEDVD-EC-----ENHLACPGQE-----CYN-----SPG 701
QY 132 TYGPDLACGG-----SORPGSGNGHCSGDGSRQDGSRCRHMGYGP 175
DB 702 SF--QCRACPSGNHLHRCRTDVDECSGARPCGPHNCT--NTESFRCSCAPGYAP 756
QY 176 L-----CTD--CMDGYF-----SLRNETHSICTA-----CDESK 204
DB 757 SGRGRCADVNECLSEGFRCFHGECINTDGSFACTCAGYRPGRGASCLDVDECSBEDL 816
QY 205 TCSGL-TNRD-----CG----- 215
DB 817 COSGICITNTDGSFECICPPGHRAGPDLASCLDVDECSERGRALCGSGRCENSPGSYRCVR 876

```

Oy 216 ECEIGW-VLDEGACVDVDECAAEPPCSAQPCKNANGSYTC-EEDSSCVCTGEBGNCN 273
Db 877 DCDPEYHAGPEGTCDVDDECQEXYGPETICGQRCENPFGSYRCTPACDPGQYPTPG--CG 933
Oy 274 CKE-----CISGY--AREHGCAADVDECSLEAKTCVRKN 305
Db 934 CQDVDECHNSRFGCAHVAQNGLSGPQCLDQGYEGARPDGRHCRVVDNDEETLQGYC--GA 991
Oy 306 ENCYNPPGSIYVCPCPDGFEETE---DACVPAEAATEGESPTQLGS 349
Db 992 ALCEVNEGSGFLCYCPCNSPEEFEDPTGRCVCPPTRSAGMFPQSGQAPA 1038

RESULT 14
O9BI05 PRELIMINARY; PRT; 2189 AA.
O9BI05
AC O9BI05;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MICRONEME PROTEIN 4.
GN MIC4.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5802;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOUGHTON:
RA Tomley F.M., Billington K.J., Bumstead J.M., Clark J.D., Monaghan P.;
RT "Emic4: A microneme protein from Eimeria tenella that contains tandem
RT arrays of epidermal growth factor-like and thrombospondin type-I
RT repeats."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ306453; CAC34726.1; -.
DR HSSP: P35555; IEMN.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF Ca.
DR InterPro: IPR000884; TSPL.
DR Pfam: PF00008; EGF_24.
DR SMART: SM00181; EGF_30.
DR SMART: SM00179; EGF_CA_30.
DR PROSITE: PS00010; ASX_HYDROXYL; 22.
DR PROSITE: PS01186; EGF_2; 18.
DR PROSITE: PS01187; EGF_CA; 18.
DR PROSITE: PS50092; TSPL_1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat
SQ SEQUENCE 2189 AA; 22438 MW; C372A420B94FCBBD CRC64;

```

Query Match	13.5%	Score 271	DB 5	Length 2189
Best Local Similarity	27.6%	Pred. No. 5e-17		
Matches	92	Conservative	24	Mismatches 87
				Indels 130
				Gaps 20

Qy	120	CVKTL--KVCCSPGTGYBDCIACOGSGRPSGNGHC-----SGDGR-----QGDG	164
Db	273	CYNLTITTFKCLCDAGYD-----AGTHESPVDIDECSEKRPNDNCNNAVCTNTEGSS	326
Qy	165	SCRCHGYG-----PLCTD--C--MDGFSSLRNETHSICT-----CG	199
Db	327	TCACKRFGSEGGFGAAGCADVDECCANSPCDAAHSCANTGCSYCTCNPGYEPASSGHC	386
Qy	200	---DE-----SCKTCSGLTNRD-----GCECEYGVLDGACVYDDECAEPPPCSAQF	246
Db	387	KDVIDECAAATACAHVSACQVNDGSEYEC--HCELEGFIDGKVCSDVDECAEAESPCCANTH	445
Qy	247	CKNANGSYCE-----PC-----DSSCVGTGE-----GG-----	272
Db	446	CLNTICISYCECKDGIGHMEGNACSIDECSEASTPEIPENCNVNTEGSEFSLAKRGYEL	505
Qy	273	-----NCKE-----CISGYARF--HGQCADVDECSLAER	299

```

Db      506  VDKGCVKIDPFARACASLHACKENPEGTAICTACTAGISGSGTATGHHDDIDEC-LAEN 564
Oy      300  TCVRKNEN--CYNTPGSYVCVCPDGEFEETDAC 330
          ++++++
Db      565  DCTPADOGICENTVGTSCKACAGYQODNSC 597

RESULT 15
088840
ID      088840      PRELIMINARY;      PRT; 3857 AA.
AC      088840;
DT      01-NOV-1998 (TREMBlrel. 08, Created)
DT      01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      MUTANT FIBRILLIN-1.
GN      FBNI.
OS      Mus musculus (Mouse).
OC      Euarystota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=B10.D2;
RX      MEDLINE=98069008; PubMed=9405934;
RA      Bona C.A., Mural C., Casares S., Kasturi K., Nishimura H., Honjo T.,
RA      Matsuda F.;
RT      "Structure of the mutant fibrillin-1 gene in the tight skin (TSK)
RT      mouse.";
RL      DNA Res. 4:267-271(1997).
DR      EMBL; AF007248; AAC62317.1; -.
DR      HSSP; P35555; IAPJ.
DR      InterPro; IPR000152; ASx_hydroxyl.
DR      InterPro; IPR002557; Chitin_binding.
DR      InterPro; IPR000561; EGF-like.
DR      InterPro; IPR001881; EGF-like.
DR      InterPro; IPR002212; TB.
DR      InterPro; IPR000822; ZnF-C2H2.
DR      Pfam; PF00008; EGF; 64.
DR      Pfam; PF00683; TB; 12.
DR      SMART; SMO0494; ChIBD2; 2.
DR      SMART; SMO0179; EGF_CA; 60.
DR      SMART; SMO0001; EGF-like; 4.
DR      PROSITE; PS00010; ASX_HYDROXYL; 61.
DR      PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR      PROSITE; PS01186; EGF_2; 50.
DR      PROSITE; PS01187; EGF_CA; 61.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR      Calcium_binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat
SQ      SEQUENCE 3857 AA; 418301 MW; 5BC0618AC527ED04; CRC64;

```

```

Query Match Similarity      13.5%   Score 270.5; DB 11; Length 3857;
Best Local Similarity      25.6%;    Pred. No.1.le-16;
Matches 104; Conservative 32; Mismatches 121; Indels 149; Gaps 24

QY      30 PCHRCRGIVDKF---NQGMDPTAKKNFGGNGTAMEEKLTSKYESSEIRLLEILGLCES 85
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      2153 PHGRCVNLIGKCYOACNPYHPTHDLF-----CVDIDECSTIMNGGET 2196

QY      86 -----SDREC-----OMLEAOEHLEAMWLIQISEYDPLEMPCVTKLKVC----GS- 129
        ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB      2197 FCTNSDGSIECCSQGFALMPDR-----SCDIDDE--CEDNPICDDGGQCTIN 2242

QY      130 -PGYGGPDCIACOG-----GSQRPCSAGNHG-----SGD-GSROGDSSCRHMGYGCP 175
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      2243 IGEYI--RCLCYDGFWASADMKTCVYNBCDLPNICLSGTCENTKGSEFIChCDMkYSgK 2300

QY      176 L---CTDMGDYFSSLRNETHSICACDESCKTGCGGLNRPDGEGEEVGMADGACAVY 211
        ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB      2301 KGRGTGTDINEBEIGAHNCNDRAVCTINTAGSR-CS-----CSPPGWIGDIGICTPL 2350

QY      232 DECAAEPPPCSAAQFCNKANGSYTC-----EBE-----DSQCVGCT 267

```

```

OY 192 THST-----CTACDESCCTCGS-----LTNDCG----- 21
Db 968 CODINECHRRPGICPDGRCVNSPASYTCLACEGEGYGGSGSCVDVNECLTPGVCAHGKCTN 10
OY 216 -----CEGVGVV-/-DEGACVVDVDECAAEPPPCSAOFCCKNANGSYTCECDSS----- 26
Db 1028 LEGSFRCSCEGGYEYTSDBEKGCDDVDECAASR-ASCPFG-LCLINTGSGFACSCAENGWVN 10
OY 263 -----CVG-CTG-EGPGNCKECISGYARE--HGACADVDECSLAERT 30
Db 1086 EDGTACEDLDCAFPAGVGVPSGCVCTNTAGSFSCDKCDGGRPSPLADSGEDVDECDPQS 11
OY 301 CVRKNEENCVNPGSVYVCVCPGPFET-----EDACVPAEAAREG 341
Db 1146 CL--GGECKNTVGSYQCLCPGFGOLANCTGVEDVNECMGEHCADPHGRLNSHG 1197

RESULT 17
ID 089103 PRELIMINARY; PRI: 644 AA.
AC 089103;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE C10/MBL/SPA RECEPTOR C10RP.
GN Lf68 OR C10RP OR AA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVEV;
RC Tenner A.J., Kim T.S.;
RA Identification of the mouse genomic DNA for C10RP."
RT Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
RN 12]
RP SEQUENCE FROM N.A.
RA Petenko O., Lemischka I.R.;
RT "Molecular characterization of AA4, an early marker of hematopoietic
RT development."
RT Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
RN 13]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RC MEDLINE=99359842; PubMed=10430665;
RX Nowsorthy P.J., Taylor P.R., Walport M.J., Botto M.;
RA "Cloning of the mouse homolog of the 126-kDa human C1q/MBL/SP-A
RT receptor, C1QRP."
RL Mamm. genome.10:789-793(1999).
DR EMBL: AF074856; AAC63274.1; -
DR EMBL: AF081789; AAC62649.1; -
DR EMBL: AF099939; AAD47906.1; -
DR EMBL: AF099938; AAD47906.1; JOINED.
DR HSSP: P35555; IEMN.
DR MGD: MGI:106664; Lf68.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_ca.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00008; EGF_5.
DR Pfam: PF00059; lectin_c.1.
DR SMART: SM00034; CLECT.1.
DR SMART: SM00179; EGF_CA.3.
DR SMART: SM00001; EGF_like.2.
DR PROSITE: PS00010; ASX_HYDROXYL.3.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01187; EGF_CA.3.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;
KW Receptor; Repeat.
SQ SEQUENCE 644 AA: 69354 MW: EB4351648BF8635A CRC64:

```

```

Query Match      13.3%; Score 266.5; DB 11; Length 644;
Best Local Similarity 27.0%; Pred. No. 3.4e-17;
Matches 11; Conservative 40; Mismatches 119; Indels 141; Gaps 28;

QY 12 LPLLLLPAPPAEAKK--ETPC-----HRCGLVDKRN-QGMVDTAKKNREGGNTAME 61
DB 143 LILDLISLTPHPSHLPKWHBSPGTPPEAPGNSLEGFLCKNFEMKCPRLALG-GPGVTTYT 201
QY 62 ----EKLKYESESEIRLELEGLC-----ESSDEFCNOMLEAOREHLEAMWLQKSEY 112
DB 202 TTPQATSSLEA--VFPASVANACGDEAKSETHYFLCN-----KT 241
QY 113 PDLFEWFCVTKLVKCCSPETYPDCLACGGSGQRPCS-GNGHCSGDSRGQDGS--CRCH 169
DB 242 PGIFHH-----SSGFLCVSPKFG-----CSFNNGCGQDCCFEGDGSFPGCGR 285
QY 170 MGYS-----GPLCTD-----CMDGTFSSLRNTHST---- 195
DB 286 PGRLLDLIVTCASRNPCSSNPCTGGMCHSVPLSENHYTCRCPSGYOLD-SSQVHCVDID 344
QY 196 -C--TACDSCCKUCGLTNRDCECEVWYLD---EGACVDVDECAAPPPCSAAGFCNK 249
DB 345 ECODSPCAQDCVNTLG--SFHC-ECWVGTOPSGPKKEACEDVDECAANSPC--AAGCIN 399
QY 250 ANGSYTCEECDSGCVGCTGEBGPNCKECISGY---AREHQCADVDECSLAE-KTCVRKN 305
DB 400 TQGSFYC-----SCKE---GYIVSGEDSTQCEDIDECSDARGNPC--D 437
QY 306 ENCYNTPGSGYVCYCPDGFETETACVPPAFAETEG-----ESTQLPSPRED 352
DB 438 SLCEFTDGSFRCGCPGWE-----LAPNGVFCSSRGTVSELPARPORKED 482

RESULT 18
Q96J98 PRELIMINARY; PRT; 2809 AA.
AC 096J98;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FIBRILLIN3.
GN KIAA1776.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
DR EMBL; AB053450; BAB47408.1;
SQ SEQUENCE 2809 AA; 300323 MW; 20C04CC006C0161F CRC64;

Query Match      13.2%; Score 264.5; DB 4; Length 2809;
Best Local Similarity 23.4%; Pred. No. 2.8e-16;
Matches 112; Conservative 37; Mismatches 146; Indels 183; Gaps 25;

QY 10 GLPLPLLLPPAPEAAKKTTPCHRCRGIVDKENQGNVDTAKKNFSGGNATWEE-----KT 64
DB 2239 GMP-----LPGSEGETDDNECAQPDLC--VNGRCVNTA-----GSFRCDCCDEGFPSPFT 2288
QY 65 LSKYSESEIR-----LLEILEGLCESSDFECNOMLEAOREHLEAMWLQKSEYDLEFEMFC 120
DB 2289 LT--ECHDIRGCPFAEVALQTMKRS-----LSSSE-----A 2318
QY 121 VTKLWCCSPGT-YGPDCLAC-----OGGSORPCSGNGHCSGDG- 158

```

```

DB 2319 VYRAECCCGGROWGPRCLCLPSTASAYRKLCPHSGSYTAGRDVDECRMLAHLCANGE 2378
QY 159 --SRODGSRCRMGX--QGPLCTDCMDYGFSSLRNTHSICNACDESKCT-----CS-- 207
DB 2379 CINSLSFRCCHCAQAGTTPATATCID-----MDECSQVPRPCTFLCKNTGSEFLCSGP 2432
QY 208 -----GLTNRDGE-----CEWVWLDGACVYDVECAAE 237
DB 2433 RGLIIEEDRTCKDLDECTSRQHNQCFLCVNTVGAFTCTCPGPFQHHQACFDNDECSAO 2492
QY 238 PPPCSAOFCKNANGSYTYEBED-----SCVGCIG---EGPNCK----- 275
DB 2493 PGPCGAHGCHHTPGSFR-CCHQGFLLVSSHGCEVNECGPAPRCQGGCONQLOGYRC 2551
QY 276 ECISGYAR--EHQGCADVDECSLAEKTCVRKNENCTNPFGYVCPDGF-----E 325
DB 2552 SCPQGFTHSQMAQVDENECALSPETC--GSAHCNNTLIGFRVCVPSGFTDQALGGCQ 2609
QY 326 TEDACV-----PPAEATEGE-----SPTQLPSREDL 353
DB 2610 DVDECAGRGRGPCSYSCANTPGFTLCGCPQGYFRAGGHCVSLGFSPPQDTPDKREL 2667

RESULT 19
Q28019 PRELIMINARY; PRT; 1963 AA.
AC 028019;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LATENT TGF-BETA BINDING PROTEIN-2.
GN LTB-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96069403; PubMed=8524260;
RA Gibson M.A., Hatzinikolas G., Davis E.C., Baker E., Sutherland G.R.,
RA Wecham R.P.;
RT "Bovine latent transforming growth factor beta 1-binding protein 2:
RT molecular cloning, identification of tissue isoforms, and
RT immunolocalization to elastin-associated microfibrils.";
RL Mol. Cell. Biol. 15:6932-6942(1995).
DR EMBL; U35363; AAA91455.1;
DR HSSP; P35355; IEMN.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR002212; TB.
DR Pfam; PF00008; EGF; 16.
DR Pfam; PF00683; TB; 4.
DR SMART; SM00179; EGF_CA; 15.
DR SMART; SM00001; EGF-like; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; 13.
DR PROSITE; PS00022; EGF_1; UNKNOWN2.
DR PROSITE; PS01186; EGF_2; 11.
DR PROSITE; PS01187; EGF_CA; 16.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 1963 AA; 211511 MW; E8E45F47FAF3B904 CRC64;

Query Match      13.2%; Score 264; DB 6; Length 1963;
Best Local Similarity 26.2%; Pred. No. 2.1e-16;
Matches 85; Conservative 21; Mismatches 84; Indels 134; Gaps 17;

QY 128 CSPG-----TYGPPCLACQGGSGORPCSGNGHCSGDSRGSGRCRMGX----- 172
DB 985 CSFGYRLHPSQAYCTDDNECL-----RDPCKGRKNCV--NRVGSYSFCYPIGYKLATS 1035

```

```

QY 173 -----OgPLCTDMDGYFSSLRNETHSI----- 195
Db 1036 GATGECDDIDCEQPGVCSRRCTNTEGSHCEDDQGYIMRKHCODINCRHPTCPD 1095
QY 196 -----CTACDE-----SC-----KTCGLNRRDCGEEVAVL 223
Db 1096 GKCVNSPSSYCLCEEBYRGCGSCVDVNECLTPGYCTHCTCINLEGSFRCSEQYEV 1155
QY 224 --DEGACVDVDECAAEPPPCSAOFCCKNANGSYNCEEDSS----- 262
Db 1156 TPDRKCKADVDECAIR-ASCTG-LCLNTBESFTCSACESYWNEDGTACEDIDCECAP 1213
QY 263 --CVG--CTG-EGPGNKECISGY---AREHGOCADVDECSLAETCYRKNENCYNTPGS 314
Db 1214 GVCSSGYCTNAGSFSCRCAGYPSALGH-TCEDVDECEDPSSCL--GGECKNTAGS 1270
QY 315 YVCYCPDGF-----ETEDACV 331
Db 1271 YQCICPPGFOLANGTVCEDVDECY 1294

RESULT 20
ID P87363 PRELIMINARY; PRT; 708 AA.
AC P87363;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE FIBRILLIN-1 (FRAGMENT).
GN FBN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RA Zhou G., Price C.E., Rosenquist T.H., Gadsen P.F., Godfrey M.,
RT "Partial cloning and sequencing of chick fibrillin-1 cDNA."
RL In Vitro Cell. Dev. Biol. Anim. 36:19-25(2000).
DR EMBL: U08872; AAB4531.1; -.
DR HSSP: P07204; 2ADX.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002212; TB.
DR InterPro: IPR001491; Thrombomodulin.
DR Pfam: PF000683; TB; 2.
DR PRINTS: PRO0907; THROMBOMODULN.
DR SMART: SM00179; EGF_CA; 14.
DR PROSITE: PS00010; ASX_HYDROXYL; 13.
DR PROSITE: PS01186; EGF_2; 10.
DR PROSITE: PS01187; EGF_CA; 13.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
FT NON_TER 708
FT SEQUENCE 708 AA; 76164 MW; C247271CIDF73361 CRC64;

Query Match 13.1%; Score 263; DB 13; Length 708;
Best Local Similarity 26.5%; Pred. No. 8.1e-17;
Matches 93; Conservative 28; Mismatches 128; Indels 102; Gaps 18;

```

```

Db 285 RLCTGDFMASEDKKTCVDVNECDLHPNICTSGTCENTKSGFICGDMGYSGKKGTTCCT 344
QY 179 DCMGYFSSLRNETHSITCTACDESCKTCGLTRNDCGCEVGVWVDEBACVYVDECAEP 238
Db 345 DINECEIHAHNCDDRAHVTNIPGSFK-CS-----CSSGIGNGCIKCYDLDCECSNGT 394
QY 239 PRCSAOFCCKNANGSYNCEEDSSCVGTGPGNCKRCISGYAREHGOCADVDECSLAE 298
Db 395 HKCSHADCKNTKMSYRC-----LCKE--GYGDGFTCTDLDCECSNL 435
QY 299 KTCYRKNENCYNTPGSYVCPDGFEEEDACVPAAEATEGSPQLPS 349
Db 436 NLC--ENQCLNAPGGRYCECDMGF-----LPSLDCKACEDIDCECLPN 477

RESULT 21
ID O35806 PRELIMINARY; PRT; 1764 AA.
AC O35806;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE LATENT TGF-BETA BINDING PROTEIN-2 LIKE PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RA Kriehn K.;
RC TISSUE=LUNG;
RT "tGF-dependent expression of a rat LTBP-2 homologue in astrocytes and
RT C6 glioma cells."
RL Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: Y12760; CA73300.1; -.
DR HSSP: P35555; 1EMN.
DR InterPro: IPR002086; Aldehyde dehydr.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002212; TB.
DR Pfam: PF00008; EGF; 16.
DR Pfam: PF00683; TB; 4.
DR SMART: SM00179; EGF_CA; 14.
DR SMART: SM00019; EGF_Like; 5.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00010; ASX_HYDROXYL; 12.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 11.
DR PROSITE: PS01187; EGF_CA; 15.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
FT SEQUENCE 1764 AA; 189866 MW; 245D57F9CE3386D0 CRC64;

Query Match 13.1%; Score 262.5; DB 11; Length 1764;
Best Local Similarity 26.0%; Pred. No. 2.6e-16;
Matches 87; Conservative 21; Mismatches 91; Indels 135; Gaps 17;

```

DB 1022 FRCSCEPEYELAPDKKCBADVEACASR-ASCP7G-LCLNTEGSEFTSCACOSGYWVEDGT 1079
 OY 262 SCVG-----CTGE-GPGNCKEVCISGYARE--HGOCADYDECSLAETCVRK 304
 DB 1080 ACEDDDECAFPGVCPTGCTNTYVGSFSCKDCDGFRRSPSLGNSCEDVDECEFGNSCL-- 1137
 OY 305 NENCYNTPGSYVCPCDFFE-----ETEDACV 331
 DB 1138 GGECKNTDGSYQCLCPGQFLANGTWCEDVDECV 1171

RESULT 22

O9WU8 PRELIMINARY; PRT; 2872 AA.
 AC O9WU8;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE FIBRILIN-1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99032689; PubMed=9815129;
 RA Kanwar Y.S., Ota K., Yang Q., Kumar A., Wada J., Kashihara N.,
 RA Peterson D.R.;
 RT "Isolation of rat fibrillin-1 cDNA and its relevance in metabolic
 RT development.";
 RL Am. J. Physiol. 275:F710-F723(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kanwar Y.S.;
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF13059; AAD34438.1; -.
 DR HSSP: P35355; IAPU.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002212; TB.
 DR InterPro: IPR000822; Znf-C2H2.
 DR Pfam: PF00008; EGF; 45.
 DR Pfam: PF00683; TB; 9.
 DR SMART: SM00179; EGF_CA; 41.
 DR SMART: SM00001; EGF_1like; 5.
 DR PROSITE: PS00010; ASX_HYDROXYL; 42.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE: PS01187; EGF_CA; 41.
 DR PROSITE: PS01187; EGF_2; 38.
 DR PROSITE: PS01187; EGF_CA; 41.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 KW SEQUENCE 2872 AA; 312069 MW; 0C4FF3B87A80280 CRC64;

Query Match 13.1%; Score 262.5; DB 11; Length 2872;
 Best Local Similarity 25.6%; Pred. No. 4.5e-16;
 Matches 104; Conservative 29; Mismatches 123; Indels 151; Gaps 23;

DB 30 PCHRCGLVDKF-----NCGVNDPAKKNGFGGNTAMEEKLTSKESSEIRLLEGLCES 85
 DB 1168 PHRCVNLIGKYECACNPGYHPHDLRF-----CYVIDECSTIMNGGCEET 1211
 OY 86 -----SDFCN-----OMLEAOEHLAWMLQKSEYPDLEFMPFCVTKLVC-----CS- 129
 DB 1212 FCTNSGSGSYSCQPGFALMPDR-----SCDIDDE--CEDNPNITDGGQCTN 1257
 OY 130 -PGTYPPDLACGG-----GSQRPCSGNGHC-----SGD-GSRQDGSRCRMHYQGP 175
 DB 1258 IPGEY--RLCYGDGFMASEDMKTCVDYNECDLNPICLSTGCTENRGSFICHCDDMGYSGR 1315
 OY 176 L-----CTDCMDGTFSSLRNETHSICACDESCCTCGSLNTRDCGECEVGVNLDGACVDY 231

DB 1316 KKGCTGCTDINECEIGHANCGRHAVCTNTAGSRK-CS-----CSPGWIGDIGIKTDL 1365
 OY 232 DECAAPPPCSAAOFCKNANGSYTC-----BEDSSCGCTGEGP----- 271
 DB 1366 DECSNETHMCSQHADCKNTMGSTRCLCKDGYTGDEFTCTDDEB--SENINLSGNQCLNA 1424
 OY 272 -----GNCX-----ECISGYA--REHG 286
 DB 1425 PAGYRCECDMGFVPSADGACEDINECSLPNICVFCTCHNLPGLFRCECEIGYELDRSG 1484
 OY 287 QCADVDECSLAETCYRKNNENCTNPGSYVCPCDFFE-----ETEDACV 331
 DB 1485 NCTDVNEC-LDPTTCI--SGNCVNTPGSYTCDDCPDPELNPTRVGV 1528

RESULT 23

O75441 PRELIMINARY; PRT; 669 AA.
 AC O75441;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 4
 DE (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98325059; PubMed=9660815;
 RA Saharinen J., Taipale J., Monni O., Keski-Oja J.;
 RT "Identification and characterization of a new latent transforming
 RT growth factor--binding protein, LTRP-4.";
 RL J. Biol. Chem. 273:18459-18469(1998).
 DR EMBL: AF054502; AAC39882.1; -.
 DR HSSP: P35355; IEMN.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002212; TB.
 DR Pfam: PF00008; EGF; 6.
 DR Pfam: PF00683; TB; 2.
 DR SMART: SM00179; EGF_CA; 6.
 DR SMART: SM00001; EGF_1like; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 6.
 DR PROSITE: PS01187; EGF_CA; 8.
 DR PROSITE: PS01187; EGF_2; UNKNOWN_2.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 FT NON_TER 1
 FT 669
 SEQUENCE 669 AA; 71702 MW; 2C76C44BB11428A4 CRC64;

Query Match 13.0%; Score 260.5; DB 4; Length 669;
 Best Local Similarity 27.3%; Pred. No. 1.3e-16;
 Matches 89; Conservative 27; Mismatches 99; Indels 111; Gaps 19;

OY 83 CESSDFECNMLEAOEHLAWMLQKSEYPDLEFMPFCVTKLVCSCPGTYPDLACOG 142
 DB 15 CEDVD-EC-----BNHLACPGQE-----CVN-----SPGSF--QCRACPS 46
 OY 143 G-----SOPRCSGNGHCSGDSRQDGSRCRMHYQGP-----CTD-- 179
 DB 47 GHHLHGRCTDYDECSGAPPCGPHHCH--NTBSFSCAPGRAPRASGRPCADVN 103
 OY 180 -CMDGYF-----SSLRNETHSICTA-----CDESKTCSGL--TNRDC 214
 DB 104 ECLGDFCFPHGECINTLTDGSACTCAPGYRPGASCLVDDECSBEDLCOSGICTNTDG 163
 OY 215 G-EC-----EYGVNLDGACVDYDECAAPPPCSAAOFCKNANGSYTCBEDSSVCGCT 267
 DB 164 SFECLCPGHRAG--PDILASCLVDDECRERGPALCSQRCENSPPSGYR----- 210

QY 268 GEGGNCRCISGY-AREHGACADVDECSLAETCYRKNCYNTPGSYVCPDGEET 326
 Db 211 -----VRDDDPGYHAGPEGICDDVNECETLQGYC--GAALCENVEGSLFCVCPNSPEEF 262
 QY 327 E---DACVPAEATEGESPTQPS 349
 Db 263 DPMTGRCVPPTTSAGMFGSQPQAPA 288

RESULT 24
 ID Q9NPY3 PRELIMINARY; PRT: 652 AA.
 AC Q9NPY3;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE D0737E23.1 (COMPLEMENT COMPONENT C10 RECEPTOR).
 GN D0737E23.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sehra H.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL118508; CAC00597.1; -.
 DR HSSP: P35555; IEMN.
 DR Interpro: IPR000152; Asx_hydroxyl.
 DR Interpro: IPR000561; EGF-like.
 DR Interpro: IPR001881; EGF-Ca.
 DR Interpro: IPR001304; lectin_c.
 DR Pfam: PF00008; EGF_5.
 DR SMART: SM00034; CLECT: 1.
 DR SMART: SM00179; EGF_CA; 3.
 DR SMART: SM00001; EGF-like; 2.
 DR PROSITE: PS00010; ASX-HYDROXYL; 3.
 DR PROSITE: PS50041; C-TYPE LECTIN_2; 1.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS01187; EGF_CA; 3.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;
 KW Receptor; Repeat.
 SQ SEQUENCE 652 AA; 68560 MW; EECADFAC55FCAC2 CRC64;

Query Match 12.88; Score 256.5; DB 4; Length 652;
 Best Local Similarity 28.08; Pred. No. 3.1e-16;
 Matches 99; Conservative 43; Mismatches 133; Indels 79; Gaps 22;

QY 3 LPRRALGLPLLLLPAA--PEAAKP-----TPCHRCGLVDPFN-OGWDTAKKNFG 54
 Db 137 ISKRCVLLLDLQPLLPRLPKWSEGPCSPGSGNIEFVCFKFGKCRPLAIG-G 195
 QY 55 GNTAMEKTLKSYSESEIRLLELEGLCESSDFECNOMLEAOEHLLEAMVQLKSEYD 114
 Db 196 PGQVTV--TTPFTTSSLEAVP---FASANVACG---EGMDDEQSHFLCKEAPD 246
 QY 115 LFEFVCYKTLKVCSPGTYPDCLACGGSGRPS-GNGHCSGGSGRODGS--CRHMG 171
 Db 247 VEDV-----SSGFLCVSPRYG---CNFNNGCHQDCFEFGDSDSLFCGCRPG 290
 QY 172 YQGLCTDCMDGYFSSLRNETHSICTADESCKTCGSLTNDGCECEVGLDGA--CV 229
 Db 291 FR--LLDDLVY---CASRNPSSSPCKRGATCYLGPHEKNTTC-RCPQGIQDSSQLDCV 344
 QY 230 DVDECAAEPPCSAAGPCCKNANGSYTCECDSSC-VGCTGEGPNCCKECLISGARHGQC 288
 Db 345 DVDEC--QDSPC--AQECVNTPGFRCF---CWVGEPEGPG-----EGAC 383
 QY 289 ADVDECSLAETCYRKNCYNTPGSYVCPDGE-----ETEDACVPP 333
 Db 384 QDVDECALGRSPCA---QGCTNTDGSFHCSCDEGYVLAGEGDQOCQDVDECVGP 434

RESULT 25
 ID Q922K8 PRELIMINARY; PRT: 685 AA.
 AC Q922K8;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE SIMILAR TO FIBULIN 1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC007140; AAH07140.1; -.
 SQ SEQUENCE 685 AA; 75283 MW; EF0D7D7F66B73B8 CRC64;

Query Match 12.88; Score 256.5; DB 11; Length 685;
 Best Local Similarity 27.58; Pred. No. 3.3e-16;
 Matches 81; Conservative 23; Mismatches 108; Indels 83; Gaps 14;

QY 115 LFEFVCYKTLKVCSPGTYPDCLACOGGS-QRP-----CSGNGHCSG 156
 Db 131 MISTQGLVFRACCVKARENSDFVONGADLQDPKIPDEEDQDPLNDRCRGGGPKQ 190
 QY 157 DGSRQDG-SCRCMGYQ---GFLCTD---CMDGYS-----SLRNETHSICTADESC 203
 Db 191 QCRGTGDEVICSCFVQGLQSDGVSCEDINCRIGSHNCRGESCINTVSGFRQGRSSC 250
 QY 204 KTGSGLTN---RDCGCEV-----GWVLDE-GACVD 230
 Db 251 GTGYELTEDNNCKDIDCEGTIHNCPDPLCQNTLGSFRCRPKLQCKSGFIQDALGNCD 310
 QY 231 VDECAAEPPCSAAGPCCKNANGSYTCECDSSCVCGCGEGSGNCKECSGY-AREHG-QC 288
 Db 311 INECLISAPCPVQQTCTINEGSYTCQK-----NVPNCGRGHLENEGRC 356
 QY 289 ADVDECSLAETCYRKNCYNTPGSYVCPDGE--FEETEDACVPAEATEG 341
 Db 357 VDVECSPPAPRC-GKGHCLNSPGSFRCCKAFYFDGISRTVDINECRRYRG 410

RESULT 26
 ID Q96HB9 PRELIMINARY; PRT: 746 AA.
 AC Q96HB9;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE SIMILAR TO LATENT TRANSFORMING GROWTH FACTOR BETA BINDING PROTEIN 3 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-COLON ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC008761; AAH08761.1; -.
 FT NON_TER 1
 SQ SEQUENCE 746 AA; 81250 MW; BD24310C7BDE18A9 CRC64;

Query Match 12.88; Score 256.5; DB 4; Length 746;
 Best Local Similarity 26.88; Pred. No. 3.6e-16;
 Matches 84; Conservative 32; Mismatches 91; Indels 107; Gaps 20;

OY	128	CSPG-----TYCPDLACGGSGRCS--GNHCSSDGSRODGSCRCMGYO-----	173
Db	91	CNPGYSHPHQRYCYVDVNEEA---EFCGGRGRTICMNTG---GSYNCHCNREGTRLHYVG	144
OY	174	GPLCTD-----CMDGYFS-----SLRNETHSICTADBSCKTCSGL	209
Db	145	GRSCVLDINECAKPHLACDGGFCINFPGHKCMCYPGYLKASRPVCDEID-CRDPSSC	203
OY	210	TNRCCGE-----CEVM-VILDEBACYVDVECAEPGPSAOFPCFNANGSYC----	256
Db	204	PDCGCKNKPGSFKCIACQPGYRSOGGCACDVNEC-AEGSPCSFG-WCENTLPSEFCTCA	261
OY	257	-----EBCDSSCV---GCTGEGPCK-ECISGY-AREHGOCADVDECS	295
Db	262	QGYPAPDPGRCLDVDCECFAGVDONCIGSNTPGSPFOCQLSGYHLRSRSHCEDIEDCD	321
OY	296	LAETCYRKNNKCNCNTGCGSVYCCPDF-----BEDDACVPRAEA-----	337
Db	322	F-PAACT--GDDCINTNGSYRLCPQHRLVLGRKCODIDEQSQDSPSLCLPHACKNLGG	378
OY	338	-----ATEGESPTQ	346
Db	379	SVCVCDEGFPTQ	392
RESULT	27		
ID	Q9NS15	PRELIMINARY;	PRT; 1242 AA.
AC	O9NS15;		
DT	01-OCT-2000 (TREMBREL_15, Created)		
DT	01-OCT-2000 (TREMBREL_15, last sequence update)		
DT	01-DEC-2001 (TREMBREL_19, last annotation update)		
DE	LATENT TRANSFORMING GROWTH FACTOR BETA BINDING PROTEIN 3.		
OS	Homo sapiens (Human).		
OC	Eumaryota; Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Shahinen J., Penttinen C., Keski-Oja J.;		
RT	"Cloning of human LTB-3,"		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; AF135960; AAF62352.2; -		
DR	HESP; P08709; IBF9.		
DR	InterPro: IPRO00152; ASX_hydroxyl.		
DR	InterPro: IPRO00561; ASX-like.		
DR	InterPro: IPRO01881; EGF_Ca.		
DR	InterPro: IPRO02212; TB.		
DR	Pfam; PF00008; EGF_13.		
DR	Pfam; PF00683; TB; 4.		
DR	SMART; SMO0179; EGF_CA; 11.		
DR	SMART; SMO0001; EGF_Like; 3.		
DR	PROSITE; PS00010; ASX_HYDROXYL; 10.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.		
DR	PROSITE; PS01186; EGF_2; 7		
DR	PROSITE; PS01187; EGF_CA; 11.		
DR	Calcium-binding, EGF-like domain; Glycoprotein; Hydroxylation; Repeat		
QW	SEQUENCE 1242 AA; 133028 MW; 155C87FB69AB221B CRC64;		

Query Match	12.88;	Score 256.5;	DB 4;	Length 1242;
Best Local Similarity	26.88;	Pred. No. 6.5e-16;		
Matches 84; Conservative	32;	Mismatches 91;	Indels 107;	Gaps 20;

QY 128 CSPG-----TYPBDCLACGGGGQRPDS -GNGHCSGDSGSGHQSGRCMHGQY----- 173
 Db 587 CNPGRSHHPQHRQYCDVYNCECA--EPCGPRGICMNTG--GSYNCHCKRGYRLHHGAG 640
 QY 174 GPLCTD-----CMDGRTS-----SLRNEHSTICTACDESCKTCSTGL 209
 Db 641 GRSQVDLNECAKPHLCGGGGGCFINPGHYKNCYPGYRLKRSAPPEVEDIDE -CRDSSC 699

```

QY      210 TNPDGE-----CEWGV-VLDEGACVDVDECAAEPPCSAAQFCNANGSYTIC---- 256
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db       700 PDGCENKPGSFCKIACQPGRYSGGGACRDVNEC-AEGSPSCFG-WCENLFGSRICTA 757
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      257 -----EECDSSCV--GCTGEGPGNCK-ECLISGY--AREHGOCADVDECS 295
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db       758 QGYAPARPGRSLDVIDEEDAGDVCGNGICSNTPGSFQQGLSGHYLSRRSHRCIDIEDCD 817
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      296 LAETCTVRKNENCYTPGSIYVCVCPDF-----ETTEDACVPAAEA----- 337
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db       818 F-PAACI--GGGCIINTNGNSYSRLCPQGHRILVGRKCQDIDECSOPSLCLPHAGACKNLGG 874
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      338 -----ATEGESPTQ 346
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db       875 SYVCVDCDEGFPTPTQ 888
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT   28
ID        Q9H7K2          PRELIMINARY;          PRT;    1382 AA.
AC        Q9H7K2;
DT        01-MAR-2001 (TREMBLrel. 16, Created)
DT        01-MAR-2001 (TREMBLrel. 16, last sequence update)
DE        01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE        FLJ00070 PROTEIN (FRAGMENT).
GN        FLJ00070.
OS        Homo sapiens (Human).
OC        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX        NCBI_TaxID=9606;
RX        [1]
RP        SEQUENCE FROM N.A.
RC        TISSUE=SPLEEN.
RA        Ohara O., Nagase T., Kikuno R., Okumura K.;
RT        "The nucleotide sequence of a long cDNA clone isolated from human
RT        spleen.";
RL        Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR        EMBL; AK024477; BAB15767.1; -.
DR        HSSP; P08709; IBP9.
DR        InterPro; IPR000152; ASX_hydroxyl.
DR        InterPro; IPR000561; EGF-like.
DR        InterPro; IPR001881; EGF_Ca.
DR        InterPro; IPR002212; TB.
DR        Pfam; PF00008; EGF; 14.
DR        Pfam; PF00683; TB; 4..
DR        SMART; SM00181; EGF; 15.
DR        SMART; SM00179; EGF_CA; 15.
DR        SMART; SM00001; EGF_like; 3.
DR        PROSITE; PS00010; ASX_HYDROXYL_1.
KW        EGF-like domain; Glycoprotein; Hydroxylation.
FT        NON_TER              1
SQ        SEQUENCE    1382 AA;  147283 MW;  AF695A4CB9913DF1 CRC64;
```

Query Match	12.8%;	Score 256.5;	DB 4;	Length 1382;
Best Local Similarity	26.8%;	Pred. No. 7.3e-16;		
Matches 84; Conservative	32;	Mismatches 91;	Indels 107;	Gaps 20;

[illegible]

Dh	718	GG-----			:	:					GACADYVNC--SEGRPSRG--MCENLPSSTYCTCAQGIKRT	758
Oy	257	-----EBCDSSCV---	-	GCTBEGPNCN-ECISGT--AREHGQCADDSCSLAEK----	299							
Dh	759	GRLSCIDVDIEEAERAKVCODGICNTICTNPESFOCCLSGHLSRDRSREDIDEDCDFFAACIG	818									
Oy	300	-TYRRKNEN-									-CYNTPGSYYCVCPDSGF	323
Dh	819	GDCINTNGSYRCPLRPLGHRLVGARKCKKDIDESCQDPGLCPHACBNLNGSYVCYCDSEGF	878									
Oy	324	EETED 328										
Dh	879	TLTDD 883										
RESULT	31											
ID	000274	PRELIMINARY;	PRT;	652	AA.							
AC	000274;											
DT	01-JUL-1997	(TEMBRel. 04, Created)										
DT	01-JUL-1997	(TEMBRel. 04, Last sequence update)										
DT	01-DEC-2001	(TEMBRel. 19, Last annotation update)										
DE	ClOR(P).											
OS	Homo sapiens (Human).											
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;											
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.											
OX	NCBI_TaxId=9606;											
RN	[1]											
RP	SEQUENCE FROM N.A.											
RX	MEDLINE-97199258; PubMed-9047234;											
RA	Nepomuceno R.R., Henschel-Edman A.H., Burgess W.H., Tenner A.J.;											
RT	"cDNA cloning and primary structure analysis of ClqR(P), the human											
RT	Clq/MBL/SPA receptor that mediates enhanced phagocytosis in vitro."											
RL	Immunity 6:119-129(1997).											
DR	EMBL: U94333; AAB5110.1; -											
DR	HSP; P35555; IEMN											
DR	InterPro: IPR000152; Asx_hydroxyl.											
DR	InterPro: IPR000561; EGF-like.											
DR	InterPro: IPR001881; EGF_Ca.											
DR	InterPro: IPR001304; lectin_c.											
DR	Pfam: PF00008; EGF_5.											
DR	Pfam: PF00059; lectin_c; 1.											
DR	SMART: SM00034; CLECT_1											
DR	SMART: SM00179; EGF_CA_3.											
DR	SMART: SM00001; EGF_Like; 2.											
DR	PROSITE: PS00010; ASX_HYDROXYL; 3.											
DR	PROSITE: PS50041; C_TYPE_LECTIN_2; 1.											
DR	PROSITE: PS01186; EGF_2; 3.											
KW	PROSITE: PS01187; EGF_CA; 3.											
DR	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat											
SO	SEQUENCE 652 AA; 68577 MW; B7EAB5FE5714A775 CRC64;											
Query Match 12.7%; Score 254.5; DB 4; Length 652;												
Best Local Similarity 30.7%; Pred. No. 4.8e-16;												
Matches 78; Conservative 27; Mismatches 88; Indels 61; Gaps												
Oy	95	EAQEHELEAMWLQLKSEYPDLFEHFCKITLVCCSPETYPDPDLAOCGGSQRPCS-GNH	153									
Dh	227	EGDDDETQSHTFLICEKAPFVFW-----GSSGPLVSPRYG---CNFNNG	270									
Oy	154	CSSGGSROGSDS--CRCHMGVGGFLCTCDMDGYFSLLRNETHSIGTACADESCSKTGSLGN	211									
Dh	271	CHQCFEBEGDGSSFICGGRPGFR--LLDDLTV---CASRNPCCSSPPRGATCYVLGPNGN	325									
Oy	212	RDCGECEYGVWLDEGA--CYVDVECAAEPPPCSAOFCCKNANGSYTCIECDSSC-VGCTG	268		</							

```

QY 312 PGSYVCVPDGFETEDACVPAEATEG-----ESPTOLPSRED 352
DB 441 DGSFRCCGCPAGFE-----LAPNGVSTCRGSMFSELPARPQRKD 479

RESULT 33
Q23587 PRELIMINARY; PRT; 3507 AA.
ID Q23587;
AC Q23587;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ZK783.1 PROTEIN.
GN ZK783.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Boufield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Shownkeen R.,
RA Smalton R., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Welstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A., Vaudin M.;
RT "The sequence of C. elegans cosmid ZK783.";
RN Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: U13646; AAC24418.1; -.
DR HSSP: P00736; IAPQ.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Trypsin.
DR InterPro: IPR001507; zona_pellucida.
DR Pfam: PF00008; EGF_15.
DR SMART: SM00179; EGF_CA_12.
DR SMART: SM00001; EGF_Like; 16.
DR SMART: SM00241; ZP_1.
DR PROSITE: PS00010; ASX_HYDROXYL; 9.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 22.
DR PROSITE: PS01187; EGF_CA; 13.
DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
KM Calcium-binding: EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SO SEQUENCE 3507 AA; 365315 MW; 154F0B687874D9DF CRC64.

Query Match 12.6%; Score 253; DB 5; Length 3507;
Best Local Similarity 19.7%; Pred. No. 4.6e-15;
Matches 103; Conservative 41; Mismatches 100; Indels 280; Gaps 23;
QY 83 CESSDFECNOMLEAQEHLNANMLQKSEYPPDLFEWFCVTLK-----125
DB 1454 CATGDHNCHEASRCQ-----NYVGACFCPTGFRKADGSCODIDECTEHN 1500

```

```

QY 126 -VCC-----SPGTGPPC-----LACGGSORPC-----SGNGHCSG----- 156
DB 1501 STCCGANAKVKNKPGTSCCECNCF LGDGYQCVPYTTKKPCDSTOSSSHCSGSESNMCEVD 1560
QY 157 -DGS-----RQDGGSCRCRNGYGP- 175
DB 1561 TVDSVECKEKMCGYKKSGKVCEDINECAKAPCSLANCVNNGNFTFSCGCKRGYGDG 1620
QY 176 -LCTD-----C--MDGYFSLRNETHS-----ICT-----ACDSCK 204
DB 1621 FMCQDINECDERHPCHPHAECTNLEG---SFKCSHGFEEDGJKCTNPLERSCDEVER 1677
QY 205 TCSGLTNRDC-----GCEVGNWLD- -GACVDNDECAEPPCS -AAQFC 248
DB 1678 FCGVNDVSVCLSVITNGSLSSVCECEPGRFFRESNCDVIDECESRNNDPASAVCV 1737
QY 249 NANGSYTC-----ECDSSCVGC-----TGEGP 271
DB 1738 NTEGSYCECAEGYEGGCVCTDIDECDRMAGCDNAMCINRMGSCGCKMAGYTGDA 1797
QY 272 -----GCK----- 275
DB 1798 TCINIEEPKSDKTACTDEMSRLCELEKQCTVDEEVPQCGACLPQHHPINTGCSLQI 1857
QY 276 -----ECI-----SGYAREHGQCADVDECSLAEKTCVKNKNCY 309
DB 1858 SGLCAQKNDCKNAECIDIDHDSHFCSGCPDGFIDGMICDDVDYCCNNA-GMCDENRKE 1916
QY 310 NTPGSYVCVPDGFETEDACVPAEATEGSEPTOLPSREDL 353
DB 1917 NTISFNCVCEGFKKVDKCYVDEKKQ-----PNEKI 1950

```

```

RESULT 34
Q9VR08 PRELIMINARY; PRT; 3680 AA.
ID Q9VR08;
AC Q9VR08;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CG15637 PROTEIN.
GN DP OR CG15637.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Aghayani A., An H.-J., Andrews-Piankovich C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cateu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster J.C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003576; AAF5100.1; -.
 DR HSSP: P35555; IEMN.
 DR Flybase: FBgn0000488; dp.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR002899; EB.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002603; ET.
 DR Pfam: PF00008; EGF_24.
 DR Pfam: PF00008; EGF_24.
 DR Prodom: PD003573; ET; 1.
 DR SMART: SM00179; EGF_CA; 16.
 DR SMART: SM00001; EGF_Like; 35.
 DR SMART: SM00274; FOLN; 5.
 DR SMART: SM00289; WRI; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 18.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 18.
 DR PROSITE: PS01187; EGF_CA; 17.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 KW KMW
 FT NON_TER
 SEQUENCE 3680 AA; 388637 MW; E5E972E1A3479EFF CRC64;

Query Match 12.6%; Score 253; DB 5; Length 3680;
 Best Local Similarity 24.6%; Pred No. 4.9e-15;
 Matches 102; Conservative 29; Mismatches 98; Indels 186; Gaps 25;

QY 82 LCSSDDECNOLEAOEHLLEAMWQ---LKSEYDLEFEMFVKTLKYC-----C-SP 130
 DB 383 LC-SNFDCTNNAECIEN--QCFCLDGFEPISGSSCVVIDE--CRTHAEVCGPHAOCLNTP 437

QY 131 GYIGPCLACQGS-----QRP-----CSGNCHSGDGRQDGSRCRHMGT----- 172
 DB 438 GSYGCECEAGYVGSPPRMACKQPCEDVRCGAHAAYCKPD--ONEAYCVGEDGTYNPSDV 494

QY 173 -----OGPLCTDCMDGYFSS---LRNETHSITACDESKT-- 205
 DB 495 AAGCVUIDECDVHGRPGSGCONATCTNSAGFTCACPFGFSDDPRISKVDVDE--CRITGA 553

QY 206 -----CSGLT-----NRDC----- 214
 DB 554 SKGAGAECEVNVPGGTYGTCRPGNTIADDPVSRCVPIVSCSANEDOPGNSICDATKRCL 613

QY 215 -----GE-----CEVGNWIDE---GACVDYDECAABP 238
 DB 614 CPEPNIGNDCRHPCEALNGAHQAQMLANGQAQCLCAPYTGTSALAGCNDIDECRAN-- 672

QY 239 PPGSAOFCFNKANGSYTC-----EEC-DSSVCGCTGEGP-----G 272
 DB 673 -FCAEKAITSNTAGYLCQPGSSSDPYRBCGCTISKYTGCSANPCANGCTCVDSYTG 731

QY 273 N--CKECLISGYAR--EHGQCADVDECSLAE-KTCVRRKNENCYTPGSIYCVCPDG 322
 DB 732 NSVC-ICRGYERHNSENGCCODVDECSYQGRKACGLNMLCKNLPSYSYECRCQG 785

RESULT 35
 Q9NTM2 PRELIMINARY; PRT; 682 AA.
 AC Q9NTM2;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE BA124N14.3 (INTRINSIC FACTOR-B12 RECEPTOR) (FRAGMENT).
 GN BA124N14.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RL Bird C.;
 RT Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 DR EMBL: AL133415; CAB87965.1; -.
 DR HSSP: P35555; IEMN.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001881; EGF_Ca.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00431; CUB; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 3.
 DR SMART: SM00001; EGF_Like; 5.
 DR PROSITE: PS00010; ASX_HYDROXYL; 4.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_4.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_2.
 DR PROSITE: PS01187; EGF_CA; 3.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Receptor; Repeat.
 FT NON_TER
 SEQUENCE 682 AA; 74166 MW; 76D118A760E6A11 CRC64;

Query Match 12.6%; Score 252; DB 4; Length 682;
 Best Local Similarity 29.1%; Pred No. 8.9e-16;
 Matches 76; Conservative 26; Mismatches 95; Indels 64; Gaps 16;

QY 118 WPCVKTLYKCSGPGTYGPDLA---CQ---GGSORPCSGNGHSGDGRQDGSRCRCHNG 171
 DB 154 FPGI-----CPROMKPGICSAADVNECEIYSGPLSCQNGTCV---NTMGSYSCHCPPE 204

QY 172 YGPFICT---DCMDGYFSSLRNETHSITACDESKTCSGLTNRDCGCEYGVWL--DE 225
 DB 205 TYGPGQASRYDDCEGG--SVARCVHGI--CEDLMREQAGEPKYSC--VCDAGWMFSPNS 257

QY 226 GAC-VVDDCAEAAPPSCSAOFCFNKANGSYTCEDCSVCGTGEAGPCKECLISGYARE 284
 DB 258 PACTLDRDECSFQPGFCSTLVOCFNTGSGFYGCACGTGNG-----NRYI-- 302

QY 285 HGOQADVDECSLAETKC-VKKNENCYTPGSIYVC-VCPDGF-----ETEDA 329
 DB 303 ---CEDINCEIINNGCSAAPVECVNTPGSSHCQACPGYGQDGRVCTLTIDICSVSNG 359

QY 330 CVPPAEATEGESF--TOLP 348
 DB 360 CHPDASCSTLGLPLCTCLP 380

RESULT 36
 Q99K58 PRELIMINARY; PRT; 1174 AA.
 AC Q99K58;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SIMILAR TO FIBULIN 2.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD. GROSS
 RC TISSUE.
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC005443; AAH05443.1; -.
 DR HSSP: P00736; IAPQ.
 DR Interpro: IPR000020; Anaphylatoxin.
 DR Interpro: IPR000152; Asx_hydroxyl.
 DR Interpro: IPR000561; EGF-like.
 DR Interpro: IPR001881; EGF_Ca.
 DR Pfam: PF01821; ANATO_2.
 DR Pfam: PF00008; EGF_6.
 DR SMART: SM00104; ANATO_3.
 DR SMART: SM00181; EGF_11.
 DR SMART: SM00179; EGF_CA_9.
 DR PROSITE: PS01177; ANAPHYLATOXIN_1; 3.
 DR PROSITE: PS01178; ANAPHYLATOXIN_2; 3.
 DR PROSITE: PS00010; ASX_HYDROXYL; 5.
 DR PROSITE: PS01186; EGF_2; 5.
 DR PROSITE: PS01187; EGF_CA; 9.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 KM SEQUENCE 1174 AA; 126460 MW; 80628AC710FBA68 CRC64;

Query Match 12.5%; Score 251.5; DB 11; Length 1174;
 Best Local Similarity 24.5%; Pred. No. 1.8e-15;
 Matches 95; Conservative 29; Mismatches 123; Indels 141; Gaps 19;

QY 83 CESS-----DFECNQL-----EAGEHELEAMMLQLS 110
 DB 525 CESPNLGYPCNHVNLSCGEDEPLIVEPRPEAPRVRSETEMASREALSIGTEA 584
 QY 111 EYP-----DLFEWCVKTL---KVCCSPGY-----GPDCLACGGSG-- 145
 DB 585 ELPLNSLGGDDDECLMPLGELCOHLCTIVGSTRACAPFEGELGGDGTCTPPDKAPOLD 644
 QY 146 -----RRCSGNGHCSGGDSRQGD--GSCRCMHGY-----QG 174
 DB 645 TAREADRESAQSPTIPLVPQPMTCCKNPGCRQVCRVGDPTAMSCSPGYAIMADG 704
 QY 175 PLCTD---CHDGYFSSLRNETHSICACDECKTCSGLTNRDCGECEYGVWLDGACVDV 231
 DB 705 VSCEDINECVTLHTCTRAE--HCVNTPGSFQC--YKALT-----CEPGYVLTGDECTDV 755
 QY 232 DECAAEPPCSAQAQFCCKNANGSYTCEEDSSCVGTGEGPNC---KECIS----- 279
 DB 756 DECVTGHNCQAGFSQNTKGSFYCQARQRCMDGFLDDPRENCYVDINCTSLLEPCRSGF 815
 QY 280 -----GY-AREHG-OCADYDECSLAETKCVRKNNENYNTPGSYVC 317
 DB 816 SCINTVGSYTCQARNPLVCGRGYHANEGSECVVNECETGVHRC--GEGQLCYNLPGSYRC 874
 QY 318 VCPDGFETEDA-----CVPPAEATG 341
 DB 875 DCKPFG--RDAGRTCIDVNECVSPG 900

RESULT 37
 008999
 ID 008999 PRELIMINARY; PRT; 1833 AA.
 AC 008999;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE LATENT TGF-BETA BINDING PROTEIN-2.
 GN LTRP2.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fang J., Li X., Smiley E., Francke U., Mecham R.P., Bonadio J.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF004874; AAB61611.1; -.
 DR HSSP: P35555; IEMN.
 DR MGD: MGI:99502; ltrp2.
 DR Interpro: IPR002086; Aldenhyde_dehydr.
 DR Interpro: IPR000152; Asx_hydroxyl.
 DR Interpro: IPR000561; EGF-like.
 DR Interpro: IPR001881; EGF_Ca.
 DR Interpro: IPR002212; TB.
 DR Pfam: PF00008; EGF_16.
 DR Pfam: PF00683; TB; 4.
 DR SMART: SM00179; EGF_CA; 16.
 DR SMART: SM00001; EGF_like; 4.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 12.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE: PS01186; EGF_2; 10.
 DR PROSITE: PS01187; EGF_CA; 16.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 KM SEQUENCE 1833 AA; 197928 MW; 3FCD6BD31E8EFC6E CRC64;

Query Match 12.5%; Score 250; DB 11; Length 1833;
 Best Local Similarity 24.8%; Pred. No. 4.3e-15;
 Matches 82; Conservative 22; Mismatches 94; Indels 132; Gaps 16;

QY 128 CSPG-----TYGPDCLACGGSGORPCSGNGHCSGDSRQGDSCRCMHGY-----QG 174
 DB 883 CSPGYQLHPGODYCTDNECM---RNPEGGRCV---NSVGSYCLCTPGITLYTLTADT 936
 QY 175 PLCTD---CHDGYFSSLRNETHSIC-----TACDESC 203
 DB 937 QECQDIDCEGPRGVCSGGRCSNTBGSYHCECDRGIMVRKHCQDINCRHPTGCPDRC 996
 QY 204 KTCGSLNRCDCGECEYGVWLDGACVDVDE-----CAEP----- 238
 DB 997 VNSPG--SYTLACBEEGVGSGSCVDVNECLTPGICTHGRCLNMEGSRCSCEPGYEV 1054
 QY 239 -----PPCSA-----AQCCKNANGSYTCEEDS-----SCVG----- 265
 DB 1055 PDKKGRDVEDCASRASPGLCLNTBESFTCSAQSGSYWNEDETACEDLDECAFPQVC 1114
 QY 266 ---CTGE--GPGNKEKICISYARE--HQACADYDECSLAETKCVRKNNENYNTPGSYVC 318
 DB 1115 PTGVCNTNVGSFSCSKDDQGGYRPNPLGNRCEDYDECBEPQSSC--RGGECKNTBGSYQCL 1172
 QY 319 CPDGFET-----EDACVPPAE 335
 DB 1173 CHQGFOLVNGTMCEDVNECVGEHCAHPGE 1202

RESULT 38
 096R09
 ID 096R09 PRELIMINARY; PRT; 3494 AA.
 AC 096R09;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE INTRINSIC FACTOR-VITAMIN B12 RECEPTOR (FRAGMENT).
 GN CUBN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Aminoff M., Brady S., Verroust P.J., Moestrup S.K., Krahe R.;

QY	226	GAG-VDVDECAAEPPPGCAAFCKANKANSYNCEEDSSVCYGTGGSPGNCRCICISGYARE	284
Db	258	PACTDTRDECSFQEPCEFTLVQCSTFYCGACPTMQG-----NGYT--	302
QY	285	HGOCADVDECSLAETC-VRRNENCYNPGSYYVC-VCPDFE-----ETEDA	329
Db	303	---CEDINECEINNGCCSVAPVECVNTPGSSHQCACPFGVGDDRVCTLTDICSVSNG	359
QY	330	CVPAAEAATEGESP--TQLP	348
Db	360	CHPDASCSTSLGSLPLCTCLP	380
<hr/>			
RESULT	40		
Q9WUH9			
ID	Q9WUH9	PRELIMINARY;	PRT; 2906 AA.
AC	Q9WUH9;		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)		
DE	FIRILLIN-2.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_Taxid=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99350231; Pubmed=10419698;		
RA	Yang Q., Ota K., Tian Y., Kumar A., Wada J., Kashihara N., Wallner E.,		
RA	Kanwar Y.S.;		
RT	"Cloning of rat fibrillin-2 cDNA and its role in branching		
RL	morphogenesis of embryonic lung.",		
Dev.	Biol. 212:229-242(1999)."		
EMBL	Af135060; Aad34439.1; -.		
HSSP	P35555; IEMN.		
DR	InterPro: IPRO02086; Aldenhyde-dehydr.		
DR	InterPro: IPRO00152; Asx_hydroxyl.		
DR	InterPro: IPRO00561; EGF-like.		
DR	InterPro: IPRO01881; EGF_Ca.		
DR	InterPro: IPRO01438; EGF-II.		
DR	InterPro: IPRO02212; TB.		
DR	pfam: PF00683; EGF; 46.		
DR	pfam: PF00683; TB; 9.		
DR	PRINTS: PRO0010; EGFBLOOD.		
DR	SMART: SMO0179; EGF_CA: 42.		
DR	SMART: SMO001; EGF_Like; 4.		
DR	PROSITE: PS00070; ALDENHYDE_DEHYDR_CYS; UNKNOWN_1.		
DR	PROSITE: PS00010; ASX_HYDROXYL; 43.		
DR	PROSITE: PS00022; EGF_1; UNKNOWN_2.		
DR	PROSITE: PS01186; EGF_2; 36.		
DR	PROSITE: PS01187; EGF_CA: 43.		
KW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.		
SO	SEQUENCE 2906 AA; 313371 MW; 9EE64E727044EF58 CRC64;		
<hr/>			
Query Match	12.4%; Score 249; DB 11; Length 2906;		
Best Local Similarity	22.0%; Pred. No. 9e-15;		
Matches	94; Conservative 54; Mismatches 121; Indels 158; Gaps 23		
<hr/>			
QY	33	RCRGLVDKF---NQGM-VDTAKKN-----FGGG--NT-AWEKTL SKYE	69
Db	1080	KCRNTIGSFKCRNGFALDMERNCTDIDECRISPDLGNGICVNTPSFCPECEGYE	1139
QY	70	SSEIILLEILE-----GLCSSD--FPC-----NQMLEAGEHEHLEAMWLQK	109
Db	1140	SGFMKKMCMQMDIDECERNPLLCRGTCVNTESFQCDCEPLGHLSPRSREDCIDINECSLS	1199
QY	110	SEYPDLFEWFMC-----VKTKLYCCSPGTG-PD-----CLACQGSGQRPCSN	151
Db	1200	DN-----LCNGKCVMNIGTYGCSCNGATATDRQCCSDIDECMINNGGCDTQCTNS	1252
QY	152	GHCSGDGSRQDGSCRCHNGY-----OGPLCTD-----CM	181

QY	DB	1253	-----	EGSYECSCSGVYALMPDGRGCADIDECENNPDI	CGGCTNINPEYRCLCY	1303
QY	182	DGYFSSLRNETHS	ICTACDESKTCSGLTNRDCGECE	-----	VGWVLDGCA	-C 228
DB	1304	DGFMA5MDMKRT	CIDYNECDLNPINICM	-----	FGECENTKGSFICHQJGYSVKKGATGC	1357
QY	229	VDVDECAEAPPPCS	AAQFCKNANGSTYC	-----	DEC	-----DSSC 263
DB	1358	TDVDECEIGAHNC	MDHMASLNPNGSPFKSCREGWNGIKICIDLD	DECACTVRKNENYNTPGSYVCYCPDG	322	
QY	264	VGCTGEGPCNCK	-EIGSYGAREHGOCADVDECSLAEKTCVRKNENYNTPGSYVCYCPDG	322		
DB	1418	VNT	-----PQSYCACSEGTGDTGDFTCSDVDEC	-AENINLCENGOCLNPGATRCCEMG	1471	
QY	323	FEETEDA	329			
DB	1472	FTRPASDS	1478			
RESULT	41					
ID	09JJS0	PRELIMINARY;	PRT;	997	AA.	
AC	09JJS0:					
DT	01-OCT-2000	(TREMBLrel. 15, Created)				
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)				
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)				
DE	CEGP1	PROTEIN.				
GN	CEGP1	OR CEGP1.				
OS	Mus musculus	(Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Bahr A., Hankeln T., Winterpacht A., Zabel B., Schmidt E.R.,					
RT	"Comparative sequencing of Human chromosome 11p15 and mouse chromosome					
RL	7."					
RL	Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.					
CC	-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.					
CC	EMBL: AJ400878; CAB92293.1; -					
DR	HSSP: P07204; IADX.					
DR	MGI: 1928765; Cegfl.					
DR	InterPro: IPR00152; Asx_hydroxyl.					
DR	InterPro: IPR000859; CUB.					
DR	InterPro: IPR00561; EGF-like.					
DR	InterPro: IPR01881; EGF-Ca.					
DR	Pfam: PF00431; CUB; 1.					
DR	Pfam: PF00008; EGF; 9.					
DR	SMART: SM00042; CUB; 1.					
DR	SMART: SM00179; EGF_CA; 6.					
DR	SMART: SM00001; EGF-like; 1.					
DR	PROSITE: PS00010; ASX_HYDROXYL; 6.					
DR	PROSITE: PS01180; CUB; 1.					
DR	PROSITE: PS01186; EGF-2; 8.					
DR	PROSITE: PS01187; EGF_CA; 6.					
KW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.					
SEQ	SEQUENCE 997 AA; 109923 MW; E0261CD9E8F70701 CRC64;					
Query Match	12.3%;	Score 247;	DB 11;	Length 997;		
Best Local Similarity	24.5%;	Pred. No. 4, 1e-15;				
Matches 119;	Conservative 34;	Mismatches 157;	Indels 176;	Gaps		
QY	6	RAALGLPLLLLP	PPRAEAKKPTFCRGRGL	-----	YDKRNGMVD	-----47
DB	10	REARALLLLLL	LLPLLAAPPD	-----	RGLTNGSEVDCAQGLDCHADALCONTR	64
QY	48	-----TAKNFGG	-----	-----	YAMEEKTLSYESSE	72
DB	65	TSYKSCSKRGY	YGEGRGCDMDRCQDNTLGGCVHDC	LNTPGNRYRC	CFPFGFALAHGHC	124
QY	73	ILLETLE	--GLCES	-----	SDFEC	-----NQ--MLEAQBHLAAMWLQKSEY 112

```

Db 125 LMDDELNNNGGCOHICTNIVGSEYECRCKGFELSDNQHICIRSEGLS----- 174
Qy 113 POLFEMECVTKLVG---CSPGTGYPDCLACOGG-----SQRPC-----SGNHCSDGS 159
Db 175 -----CMKDKGCGHICKEAPRGSVACBRCRPFELANNQKPCILTCNHNNGGCCQHSCE 227
Qy 160 RQDGG-SCRCHMGYQ---GPLC-----TDCMDGYFSSLNETHSICT-- 197
Db 228 DTFNEGPECSCHPRYRLHADRSCLDEQGYLLEGTESNATSVADGDKRVKRRLLMECAVN 287
Qy 198 --ACDSCKTCSGLTNRDGCEYGVNL--DEGACVYDVDECAAEPPPCSAOFCCKNANGS 253
Db 288 NGGCDRTCKDTS--TGVMHC-SCPTGFTLQYDGKTKDIDECQTRNGGCN--HFCKNTVGS 342
Qy 254 YTC-----EECDSSCVGTGCGEPGNCCKECISJGAREHGQCA 289
Db 343 FDCSCKKGRFLTDEKSCQDVDECSLERTCDHSCINHPTFCACNGYTLYSFTH--CG 400
Qy 290 DVDCESLAEKTCVRKNENCYNTPGSYVVCPPDGF--ETEDACVPPAEAEATGESPTQL 347
Db 401 DTFNEGSVNNNGC---QOVCTINTVGSYECQCHPGFKLHMNKKDCV-----EVAGFPPTSM 451
Qy 348 PSREDL 353
Db 452 TPRVSL 457

```

```

RESULT 42
Q9HB05 PRELIMINARY; PRT: 495 AA.
ID Q9HB05;
AC Q9HB05;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 54.3 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF217999; FAG17241.1; -.
DR HSSP: P35355; 1EMN.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PR00008; EGF; 6.
DR SMART: SM00181; EGF; 9.
DR SMART: SM00179; EGF_CA; 9.
DR SMART: SM00001; EGF-like; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
KW EGF-like domain; glycoprotein; Hydroxylation; Hypothetical protein.
SQ SEQUENCE 495 AA; 54340 MW; C40434B6C82E3D70 CRC64;

```

Query Match 12.3%; Score 246.5; DB 4; Length 495;
 Best Local Similarity 28.4%; Pred. No. 2, 1e-15;
 Matches 83; Conservative 23; Mismatches 97; Indels 89; Gaps 17;

```

Qy 120 CVYTLKVC--SPGTGYPDCLACOGGSRP-----CGSNHCSDGS 159
Db 6 CGGVFRACCCVKSQETGDLVY---GGIQETDKLIEVEEEDPYLNDRCGGGPKQOCR 61
Qy 160 RQDGG-SCRCHMGYQ---GPLCTD---CMDGYFS-----SLNETHSIC--- 196
Db 62 DTGDEYVVCSCFVGYQLLSDVSCEDYNECTIGSHSLGESCINTVGSFRCQDSSCGTG 121

```

```

Qy 197 --TACDESKC---TC-SGLTN-----RDCGECGWLDE-GACVYDDE 233
Db 122 YELTEDNSCKDDICEGSGIHNCPLPDFICQNTLGSFRCRKLQCKSGFIQDALGNCIDNE 181
Qy 234 CAAEPPPCSAOFCCKNANGSYTCEEDSSCVGTGEGPNCKEICISGY-AREHG-QCAVY 291
Db 182 CLTISAPCRPGHCTINTGESYTCOK-----NVPNCGRGYHINEGTCVDPV 227
Qy 292 DECSLAEKTCVRKNENCYNTPGSYVVCPPDGF--FEETEDACVPPAEAEATG 341
Db 228 DECAPPAEPC-GKGRVCVNSPSSFRCECKTGYTFDEISRMCAVDYNECQRRYP 278

```

```

RESULT 43
Q20911 PRELIMINARY; PRT: 3871 AA.
ID Q20911; Q23242;
AC Q20911; Q23242;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ZC116.3 PROTEIN.
GN ZC116.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Faveille A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan R., Waterston R.,
RA Watson A., Weinstock K., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 18 CUB DOMAINS.
DR EMBL: Z74473; CAA98952.1; -.
DR EMBL: Z74046; CAA98952.1; JOINED.
DR EMBL: Z74473; CAA98557.1; JOINED.
DR HSSP: P00740; 1EDM.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF_CA.
DR Pfam: PF00431; CUB; 13.
DR Pfam: PR00008; EGF; 7.
DR SMART: SM00042; CUB; 18.
DR SMART: SM00179; EGF_CA; 3.
DR SMART: SM00001; EGF-like; 5.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 20.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_4.
DR PROSITE: PS01186; EGF-2; 2.

```

DR PROSITE; PS01187; EGF_CA; 3.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
 SO SEQUENCE 3871 AA; 433663 MW; 843600647421AB12 CRC64;

Query Match 12.2%; Score 245.5; DB 5; Length 3871;
 Best Local Similarity 28.8%; Pred. No. 2,7e-14;
 Matches 74; Conservative 26; Mismatches 94; Indels 63; Gaps 15;

OY 128 CSQCTVPCDLA-----CG--GSGRPPCSGNGHCSGDSRGDSCGRCHMGYQGFLLTCDM 181
 DB 100 CPPEFTGTTEADIDECSSVYNGTTAGQNNCTCI--NNNGGFECCQCSGHHGLCO--- 153
 OY 182 DGYFSLRNETHSICTA---CDESCKTCSGLTNHDCE-----CEGVWLV-----DEGACVD 230
 DB 154 --YHMAKSGSTFELCGPHGHCIESIVPTGQSSDITTYKICIDMGKRVSSDKNNPTCYD 211
 OY 231 VDECAEPSPCSAAQFCNNANGSYTCECDSS-----CVG---CTGE----- 269
 DB 212 VNEC--ESNCHPGVDCINLPGSFVSCGPKGYKTDGNVICIDVNECGEIRVCSPLSKH 269
 OY 270 ---GPGNCKCISGYAHBHQCADVDEC-----SLAEKTCVRKNENCMYTPGSYVCP 320
 DB 270 NTLGSIYCDSCPTGYSGDGGNCVAKDSCVXNKCHKLA--TC-KVTDGYSAGVDYTCYCP 326
 OY 321 DGEFEET---EDACVPPA 334
 DB 327 DGYVGDGIGEGCVKSA 343

RESULT 44

Q25678 ID Q25678 PRELIMINARY; PRT; 733 AA.

AC 025678;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE FIBRILLIN (FRAGMENT).
 OS Podocoryne carnea.
 OC Eukaryota; Metazoa; Chordata; Hydrozoa; Hydroids; Anthomedusae;
 OC Hydroclitidae; Podocoryne.
 OX NCBI_TaxID=6096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95301106; PubMed-7781906;
 RA Reber-Muller S., Spissinger T., Schuchert P., Spring J., Schmid V.;
 RT "An extracellular matrix protein of jellyfish homologous to mammalian
 RT fibrillins forms different fibrils depending on the life stage of the
 RT animal";
 RL Dev. Biol. 169:662-672(1995).
 DR EMBL; L39930; AAA91336.1; -;
 DR HSSP; P35555; 1EMN
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001128; Cyt_P450.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_CA.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR002212; TB.
 DR InterPro: IPR001491; Thrombomodulin.
 DR Pfam; PF00008; EGF; 14.
 DR PRINTS; PRO0010; EGFBL00D.
 DR PRINTS; PRO0907; THROMBOMODULN.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00001; EGF_like; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 12.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE; PS01186; EGF_2; 11.
 DR PROSITE; PS01187; EGF_CA; 10.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 FT NON_TER 1 1
 FT NON_TER 733 733
 SO SEQUENCE 733 AA; 79856 MW; D65699052731D5E7 CRC64;

Query Match 12.2%; Score 244; DB 5; Length 733;
 Best Local Similarity 25.1%; Pred. No. 5.7e-15;
 Matches 92; Conservative 24; Mismatches 85; Indels 166; Gaps 21;

OY 111 EYPLFEMFVCVKLVKCCSPGT-----YPCDLAC-----OGSQRPPCSGNGHCSGDG--- 158
 DB 171 EYPL-----CRHGSCSNKIGSFMCQCNQEGFKDQATNQSCODINECQDPCK 218
 OY 159 -----SRQDGSRCRCHMGYQ---GPLCTD-----CMD 182
 DB 219 NGNCRNRIGSAVCTCYGVEKTEIDGLSCDEKDECADENRQGLGTCVNTDGFKLCNP 278
 OY 183 GY-----TSSLRNE-----THSICTAD-----ESKTC--- 206
 DB 279 GYVSDENRECIDVRSFCSLSLENNOCNTRANGLNVTKSYV-CCSMGAGFDDPELCVVK 337
 OY 207 -----SGL-----TNRDGECEVG-----WVLDEG----- 226
 DB 338 GTRFEYCYCPNGVGFITSNNTDINECANLGTCLNGCQVNAVDSFRCVCDGTYLNRNP 397
 OY 227 -ACVDVDECAAEPPPCSAQFCNNANGSYTCECDSSCVGCTGEPGNCREKISGYAREH 285
 DB 398 RECIDRECAAGSNYCGNG-CTNLVGSYOC-SCEE-----GPEPT-----DS 439
 OY 286 GQCADVDECSLAETCYVRKNENCMYTPGSYVCPDGFEEEDA--CYPAEAETAGES 343
 DB 440 PACVDVDECAKELCYRCKNLY--GSYMCSCPKGFKLAEDQKCV-----DINCDT 491
 OY 344 PTOLPSR 350
 DB 492 PHRCPR 498

RESULT 45

Q9NQ36 ID Q9NQ36 PRELIMINARY; PRT; 999 AA.

AC Q9NQ36;
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE CEGP1 PROTEIN.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bahr A., Hankeln T., Winterpacht A., Zabel B., Schmidt E.R.;
 RT "Comparative sequencing of human chromosome 11p15 and mouse chromosome
 RT 7.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL; A1400877; CAB92285.1; -;
 DR HSSP; P07204; 1ADX.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_CA.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00008; EGF; 5.
 DR SMART; SM00042; CUB; 1.
 DR SMART; SM00179; EGF_CA; 6.
 DR SMART; SM00001; EGF_like; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 6.
 DR PROSITE; PS01180; CUB; 1.
 DR PROSITE; PS01186; EGF_2; 8.
 DR PROSITE; PS01187; EGF_CA; 6.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SO SEQUENCE 999 AA; 109956 MW; 61334844A0053095 CRC64;

Query Match 12.2%; Score 244; DB 4; Length 999;
 Best Local Similarity 25.1%; Pred. No. 8.1e-15;
 Matches 120; Conservative 31; Mismatches 166; Indels 162; Gaps 28;

```

QY 7 AALGLPLLLLPAPAAKPPCH-RCRGL--VDKFNQGVYPAKKNFGGNTAMBE 62
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11 AAMAVLLLLLPPLLLLAGAVPPGRRAAGPQEDVDECAOGL-DDCHADALCQNTPTSY 69
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 KTLISK--YSSSEIRLEILEG--LCSSDFECNQMLAEQEHLEAMWLOKSEXP-DLFEEM 118
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 70 KSCSKPGTGG-----EGROCEDID-ECGNELNGCVH--DCLNTPGNTRCTCFDG 116
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 119 F-----CYKTL--KYCCSPGY-----GPD 136
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 FMLAHGHNCLDDECLENNNGGQHTCVANMGSECCKEGFLSDNQHCHIRSEGLS 176
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 CL-----ACO-----GGSQRPC-----SGNHCSDGSRQGDG-S 165
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 CMNKHGCSHICKKAPRGVACECRPGFELAKNORCIIITCNHNGSCQHSQCDTADGPE 236
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 CRCHMGYO---GPLCTDCMDGYFSSLRNETHSI-----CT-----ACDPS 202
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 CSHPPYKMHHTDRSCLEREDIVLEVTESVTVYDGDKRVKRLMETCAVANNNGCDDRT 296
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 203 CKTCSGLTRNDGCECEVGVW--LDEGACVYDDECAAEPPPCSAOFCCKNANGSYTC----- 256
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 CKDTS--TGVMC-SCPVGFLQIDGKTKDIDECQTRNGGCD--HFCKNIVSGFDGCKK 351
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 257 -----EECDSSCVGCTGCGPNCKECISGVARHGQCADVDECSL 296
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 GFKLLTDEKSCQDVEDCSLDRCTDHSICINHPGFACACNNGYLYGPTH--CGDTRECSI 409
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 297 AEKTCVRKNKNCYNTGPGSVYCVCPDGEF--ETFDACVPAEAATGESPOTOPSRNEDL 353
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 NNGGC---QQVCVNTVSGTECGCHPGYKILHMKKDCV-----EVKGLPLTSPVPSVL 459
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 46
 091VK0 PRELIMINARY; PRT: 915 AA.

Query Match 12.1%; Score 243; DB 11; Length 915;
 Best Local Similarity 29.5%; Pred. No. 9.1e-15;
 Matches 65; Conservative 21; Mismatches 86; Indels 49; Gaps 13;

```

QY 125 KVC---CSPCTYGPDLACOGSGQSPCSGNGHSGDGSROGDSRCRCHMGY-----QGPL 176
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 670 RIVSSCPREHYADKKRC-----RKCAPP--CESCFGSIGNCLSKYIFLINEETSSC 722
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 177 CTDCMDGYFSSLRNETHSICTACDESKCTCSGLTRNDGCECEVGVWLDGACVYDDECA 236
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 723 VTCQCPGYSYEDIK---NWCCKSENCACIGPHN--CTECKGGLSLQSGRC----- 769
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 237 EPPPCSAOFCCKNANGSYTCEDDSSCVGCTGCGPNCKECISGVARHGQCADVDECSL 296
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 770 -SVTCEDGQF---FNG-HDQCPHRCATCGAGADGCTNCTGYVMEBGRG--VQSCSV 822
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 297 A-----EKTCVRKNKNCY--NTPGSYVC-VCPDGF 323
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 823 SYLLDHSSEGGYKSKCKRCNDSCLTCNGRPFKNCSQSPSY 862
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 47
091V88 PRELIMINARY; PRT: 561 AA.
AC 091V88:
DB 01-DEC-2001 (TREMBLrel. 19, Created)
DB 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DB 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POEM (NEPHROTECTIN SHORT ISOFORM).
GN POEM OR NEPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21551216; PubMed=11546798;
RA Morimura N., Tezuka Y., Watanabe N., Yasuda M., Miyatani S.,
RA Hozumi N., Tezuka K.;
RT "Molecular cloning of POEM. A novel adhesion molecule that interacts
  with alpha2beta1 integrin."
RL J. Biol. Chem. 276:42172-42181(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH SWISS; TISSUE=KIDNEY;
RX MEDLINE=21363579; PubMed=11470831;
RA Bradenberger R., Schmidt A., Linton J., Wang D., Backus C., Denda S.,
RA Muller U., Reichardt L.F.;
RT "Identification and characterization of a novel extracellular matrix
  protein nephronectin that is associated with integrin alpha2beta1 in
  the embryonic kidney."
RL J. Cell Biol. 154:447-458(2001).
DR EMBL: AB059656; BAB69692.1;
DR EMBL: AY035898; AAK96010.1;
SQ SEQUENCE 561 AA; 61490 MW; 69E7ACAADEE3F506 CRC64;

```

Query Match 12.0%; Score 241.5; DB 11; Length 561;
 Best Local Similarity 25.3%; Pred. No. 7.3e-15;
 Matches 74; Conservative 25; Mismatches 113; Indels 81; Gaps 12;

```

QY 105 WLQKSEYDLEFEPFCVTKLVKCCSPGYGP--DCL-----ACGGSGRPPCSGNCH 153
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14 YLOVADFGRMPROIYSSIGLC---RYGRIIDCWMGAROSWGQCPVCOPOCK-HGE 68
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 154 CSGDGSROGDSRCRCHMGYQGLCTD-----CMDGYF-- 185
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 69 C-----VPPNCKCKHPGAGKTCNODLNEGCIKPPCKNHRQMTFGSKYCYCLMGMILL 122
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 186 --SSLNETHSICTACDESKCTCSGLTRNDGCECEVGVWLDGACVYDDECAEPPPCSA 243
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 PDGSSSALSCSMANQYQCDVYKGVRCQPSGLQLPDRTRCYDVIDECATGRVSCR 182
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 ADFCKNANGSYTCEDDSSCVGCTGCGPNCKECISG---YAREHGQCADVDECSLAEK 299
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 FRQCVTFGYSYIC-----KCHTGFPLMYIGKYQCHDIDECSLGQH 223
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 300 TCVRKNCNYPNGSVYCVCPDGEFETEDACV--PPAEATEGSPGLPSR 350
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 224 QC--SSTARCYNIHGSYKCCQCRDGYEGDGLNCVYIPKVMTEPS--GPIHMPER 272
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 48
 090H16 PRELIMINARY; PRT: 554 AA.
 ID 090H16
 AC 090H16;

DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE ENDOPROTEASE FURIN.
GN FURIN.
OS Spodoptera frugiperda (fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxId=7108;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SF9, FALL ARMYWORM OVARY;
RA Clepik M., Klenk H.;
RT "Cloning and functional characterization of FURIN from Spodoptera
RT frugiperda (Sf9) cells."
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z68888; CAA93116.1; -.
DR HSSP; Q99405; IMPT.
DR MEROPS; S08.0PB; -.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR002884; P_domain.
DR Pfam; PF01483; P; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00261; FU; 10.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 1299 AA; 142020 MW; 4C3799C7BBC572AB CRC64;

Query Match 11.9%; Score 239; DB 5; Length 1299;
Best Local Similarity 23.8%; Pred. No. 3.3e-14;
Matches 92; Conservative 32; Mismatches 146; Indels 86; Gaps 20;
QY 19 PPAPEAAKKPTPCRCRGVLDKFNQGNVDPAKKNFGGNTAWMEKTLSS--KYE-SSEIR 74
DB 881 PPAYYADKKRKECMRCP-----VCSCTCTSAFCLSCPEKMLNKKK 922
QY 75 LLEILEGLCESSDF-----ECNOMLEA---QEEHLEAMWLQLKSEYPLDF-EMFCYK 122
DB 923 CMFVGSDDKCSAGFAVDQCKRCNPACDSCYGENEGHC-----LTCPPNPLLDYKCV 976
QY 123 TLKVCSPGYGP--DCLACQGGSGRPGCSGNHCSGDGSGRQDSCRGCHMGYQPLC-ID 179
DB 977 E---CSKGYAEAGRCARCMHG-----CSDCVSRNLCTSCASTLRLOSGACRTS 1022
QY 180 CMDGYFSSLNETHSICITACDESCIKTCGILTRNDGCEGEYGVLDGACVDECAEAPP 239
DB 1023 CADGYTAD-----RGTSKCYLSCRTCTIGPRRQOCASCPBGWRLAAGEC-----HP 1068
QY 240 PCSAAQFCCKNANGSYTCECDSSCVGCTGEGPNCKECISGYAREHGQCAD-----VDE 293
DB 1069 ECPQG--FYGSPG---CRHCHHCRCDGSGPLHCKSCPPRFMLDGLCMECIGSQYIDA 1124
QY 294 CSLAEKTCYAKNKNCNTPGSYVC--VC--PDGFEETEDACVP--PARAEATGESPT 345
DB 1125 TSGTCRSCDASCRTG--SGPQGFSTCTCSRLRLIDRLNOCVPCCSBGRGVNSTPPT 1179

Search completed: September 10, 2002, 11:13:37
Job time: 234 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 11:09:23 : Search time 13.41 seconds
(without alignments)
1019.239 Million cell updates/sec

Title: US-09-905-075-2

Perfect score: 2005

Sequence: 1 MRLPRRAIGLEPLLPLP.....AEATEGESEPTQLPSREDL 353

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt-40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	286.5	14.3	684 1	FBL1_CHICK
2	274.5	13.7	2871 1	FBN1_PIG
3	270	13.4	2871 1	FBN1_BOVIN
4	268.5	13.4	2871 1	FBN1_HUMAN
5	263	13.1	1877 1	PKC5_MOUSE
6	254.5	12.7	703 1	FBL1_HUMAN
7	253.5	12.6	705 1	FBL1_MOUSE
8	252.5	12.6	2871 1	FBN1_MOUSE
9	252.5	12.6	2907 1	FBN2_MOUSE
10	249.5	12.4	2911 1	FBN2_HUMAN
11	248.5	12.4	1696 1	PKC5_BRACL
12	247	12.3	675 1	PRTS_RAT
13	247	12.3	1221 1	FBL2_MOUSE
14	242.5	12.1	1184 1	FBL2_HUMAN
15	240.5	12.0	1394 1	TGFB_HUMAN
16	239	11.9	676 1	PRTS_HUMAN
17	239	11.8	1877 1	PKC5_RAT
18	236	11.6	913 1	PKC5_HUMAN
19	232	11.6	675 1	PRTS_MOUSE
20	232	11.6	1375 1	NID2_HUMAN
21	228.5	11.4	649 1	PRTS_MACMU
22	228	11.4	1680 1	FUR2_DROME
23	226	11.3	2531 1	NTC1_MOUSE
24	225.5	11.2	646 1	PRTS_RABIT
25	224.5	11.2	1403 1	NID2_MOUSE
26	223	11.1	2524 1	NOTC_XENLA
27	222.5	11.1	886 1	EMRI_HUMAN
28	222	11.1	675 1	PRTS_BOVIN
29	222	11.1	1712 1	TGFB_RAT
30	222	11.1	2444 1	NTC1_HUMAN
31	221	11.0	1247 1	NID2_HUMAN
32	220	11.0	1247 1	NTC1_RAT
33	217.5	10.8	3672 1	LMT2_CAEEL

34	216.5	10.8	2703 1	NOTC_DROME	P07207 drosophila
35	216	10.8	448 1	FBL5_RAT	O9wv8 ratu muscu
36	214.5	10.7	448 1	FBL5_MOUSE	O9wv9 ratu muscu
37	213.5	10.6	1245 1	NID2_MOUSE	P10493 mus muscu
38	212	10.6	2318 1	NTC3_MOUSE	O61962 mus muscu
39	211.5	10.5	810 1	NEL1_HUMAN	O92832 homo sapien
40	211.5	10.5	931 1	EMRI_MOUSE	O61549 mus muscu
41	211.5	10.5	2437 1	NOTC_BRARE	O61530 brachydanio
42	209.5	10.4	448 1	FBL5_HUMAN	O9ubx5 homo sapien
43	209.5	10.4	687 1	VS41_GIALA	P92127 giardia lam
44	209	10.4	2139 1	CRB_DROME	P10040 drosophila
45	208	10.4	443 1	FBL4_CRIGR	O55058 cricetus
46	207	10.3	443 1	FBL4_MOUSE	O9wv9 mus muscu
47	205.5	10.2	473 1	FP2_MYGA	O25464 mytilus gail
48	204.5	10.2	810 1	NEL1_RAT	O62919 ratu muscu
49	204.5	10.2	1964 1	NTC4_MOUSE	P31695 mus muscu
50	203	10.1	443 1	FBL4_HUMAN	O95967 homo sapien
51	202.5	10.1	356 1	TRBM_BOVIN	O06579 bos taurus
52	202.5	10.1	712 1	FBL1_CAEEL	O77469 caenorhabdl
53	202.5	10.1	816 1	NEL2_RAT	O62918 ratu muscu
54	202	10.1	667 1	TS11_GIALA	O03185 giardia lam
55	201.5	10.0	816 1	NEL2_MOUSE	O61220 mus muscu
56	201	10.0	383 1	DLK_HUMAN	P80370 homo sapien
57	201	10.0	433 1	FBL3_HUMAN	O12805 homo sapien
58	201	10.0	835 1	CD97_HUMAN	P48960 homo sapien
59	201	10.0	4289 1	TENX_HUMAN	P22105 homo sapien
60	200.5	10.0	1064 1	FBL1_STRPU	P10079 strongylocet
61	200.5	10.0	4655 1	LRP2_HUMAN	P98164 homo sapien
62	200	10.0	713 1	TSB4_GIALA	P21849 giardia lam
63	197	9.8	577 1	TRBM_MOUSE	P15306 mus muscu
64	197	9.8	723 1	DL11_HUMAN	O00548 homo sapien
65	196.5	9.8	816 1	NEL2_HUMAN	O99435 homo sapien

ALIGNMENTS

RESULT 1	ID	FBLL_CHICK	STANDARD:	PRT:	684 AA.
AC	073775;	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Fibulin-1 precursor.				
GN	FBLN1.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OX	Gallus.				
RN	NCBI_TaxID=9031.				
RP	SEQUENCE FROM N.A.				
RA	Barth J.L., Argaves K.M., Roark E.F., Little C.D., Argaves W.S.;				
RT	Isolation of chicken and nematode fibulin-1 homologs and				
RT	characterization of the nematode fibulin-1 gene.*;				
RL	Submitted (FE8-1998) to the EMBL/GenBank/DBJ databases.				
CC	- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.				
CC	- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.				
CC	- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL: AF051400; AAC05388.1; -				
DR	HSSP: P00742; IHCG.				
DR	InterPro: IPR000020; Anaphylatoxin.				
DR	InterPro: IPR000152; Asx_hydroxyl.				

DR	IProPro:	IPRO00561;	EGF-Like-
DR	InterPro:	IPR001881;	EGF_Ca.
DR	Pfam:	PF01821;	ANATO_2.
DR	Pfam:	PF00008;	EGF; 6.
DR	SMART:	SM00104;	ANATO; 3.
DR	SMART:	SM00179;	EGF_CA; 8.
DR	SMART:	SM00001;	EGF_Like; 1.
DR	PROSITE:	PS00010;	ASX_HYDROXYL; 5.
DR	PROSITE:	PS01177;	ANAPHYLATOXIN_1; 1.
DR	PROSITE:	PS01178;	ANAPHYLATOXIN_2; 2.
DR	PROSITE:	PS00022;	EGF_1; FALSE_NEG.
DR	PROSITE:	PS01866;	EGF_2; 3.
DR	PROSITE:	PS01187;	EGF_CA; 8.
KW	Signal:	Glycoprotein;	Extracellular matrix; Repeat; EGF-like domain;
KW	Calcium-binding.		
FT	SIGNAL	1..25	POTENTIAL.
FT	CHAIN	26..684	FIBULIN-1.
FT	DOMAIN	33..74	ANAPHYLATOXIN-LIKE 1.
FT	DOMAIN	75..109	ANAPHYLATOXIN-LIKE 2.
FT	DOMAIN	110..142	ANAPHYLATOXIN-LIKE 3.
FT	DOMAIN	177..216	EGF-LIKE 1.
FT	DOMAIN	217..262	EGF-LIKE 2.
FT	DOMAIN	263..308	EGF-LIKE 3.
FT	DOMAIN	309..356	EGF-LIKE 4.
FT	DOMAIN	357..399	EGF-LIKE 5.
FT	DOMAIN	400..441	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	442..481	EGF-LIKE 6.
FT	DOMAIN	482..525	EGF-LIKE 7.
FT	DOMAIN	526..570	CALCIUM-BINDING (POTENTIAL)
FT	DISULFD	33..39	BY SIMILARITY.
FT	DISULFD	34..66	BY SIMILARITY.
FT	DISULFD	47..67	BY SIMILARITY.
FT	DISULFD	76..107	BY SIMILARITY.
FT	DISULFD	89..108	BY SIMILARITY.
FT	DISULFD	110..134	BY SIMILARITY.
FT	DISULFD	111..141	BY SIMILARITY.
FT	DISULFD	124..142	BY SIMILARITY.
FT	DISULFD	181..191	BY SIMILARITY.
FT	DISULFD	187..200	BY SIMILARITY.
FT	DISULFD	202..215	BY SIMILARITY.
FT	DISULFD	221..234	BY SIMILARITY.
FT	DISULFD	228..243	BY SIMILARITY.
FT	DISULFD	249..261	BY SIMILARITY.
FT	DISULFD	267..280	BY SIMILARITY.
FT	DISULFD	274..289	BY SIMILARITY.
FT	DISULFD	295..307	BY SIMILARITY.
FT	DISULFD	313..326	BY SIMILARITY.
FT	DISULFD	320..335	BY SIMILARITY.
FT	DISULFD	342..355	BY SIMILARITY.
FT	DISULFD	361..374	BY SIMILARITY.
FT	DISULFD	368..383	BY SIMILARITY.
FT	DISULFD	385..398	BY SIMILARITY.
FT	DISULFD	404..416	BY SIMILARITY.
FT	DISULFD	412..425	BY SIMILARITY.
FT	DISULFD	427..440	BY SIMILARITY.
FT	DISULFD	446..455	BY SIMILARITY.
FT	DISULFD	451..464	BY SIMILARITY.
FT	DISULFD	466..480	BY SIMILARITY.
FT	DISULFD	486..499	BY SIMILARITY.
FT	DISULFD	495..508	BY SIMILARITY.
FT	DISULFD	510..524	BY SIMILARITY.
FT	DISULFD	530..543	BY SIMILARITY.
FT	DISULFD	537..552	BY SIMILARITY.
FT	DISULFD	557..569	BY SIMILARITY.
FT	CARBOHYD	96..96	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	536..536	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	540..540	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	638..638	N-LINKED (GLCNAC. . .) (POTENTIAL).
QO	SEQUENCE	684 AA; 75623 MW; 1638DBA094739199 CRC64;	

Query Match	14.38;	Score 286.5;	DB 1;	Length 684;
Best Local Similarity	25.78;	Pred. No. 2.8e-14;		

Matches	127;	Conservative	37;	Mismatches	120;	Indels	211;	Gaps	26
QY	6	RALGLLPLLLPLPPAPFAKRPTECHCRGLVDKFNQGMVDTAKNNFGGNTAMEETL							65
Db	5	RRARPLRLLLLLALLPLALRGDLSMEBC---	CDK---						44
QY	66	SKYESSEIFLLEILEGLESSDPECNOMLEAO-----	EEHLEAMVLLKSEPYDL----						115
Db	45	-----RICHSLPLISSR--ECS-WIYQYCCRSKLEEHNCSDIGIFASVHECDSDHN							93
QY	116	-----EEMF-----	CVKTLKV-----						138
Db	94	GENSTCEAEYEFKRCYCCLLGKTAQVQOSEPNKIKYOGGIYFRACCIVVGOEGTDVSI							153
QY	139	-----	AGCGSQRCRSGNGHCSGDGSGRQDGSCRHMGY						172
Db	154	SDDARKKEVEISKELLQDEPYLHDGCRGG--PCS--QCCRDITGSSY--VCSCEVGY							206
QY	173	Q-----GPIQTD---CMDGYFSLRNETHS--ICTAC-----	DESCRTCSGLTN--						211
Db	207	QLQPDGVNCEIDINECITG-----	THSCGIGQTCVMTLGSFRCORPDTSCGYELHDD						258
QY	212	---RCCGEV-----	GWVLDE-GACVYVDECAERP						239
Db	259	SRCKIDIEETGTNHCPPDIFQNTPGSFRCRPRKLQCMNGFIQDALGNICIDINECLSTNM							318
QY	240	PCSADQFCNKANGSYTC-----	EECDSSCGVGTGE-----						270
Db	319	PCPAGQICINTDGSVTCQRISPSCRGYHLNEDGRVADYDCCSSDQPC-GEHVACING							377
QY	271	PGNCK-ECISGYARE--HGQCADVDEC-SLAETKYVRKNENCYNTPGSYVYVCPGPFEEY							326
Db	378	PENYCECKSGSGSFVYSITRKTIDINECRYPGRICAHK--CENTPGSYCTCTTGFIKS							434
QY	327	ED--ACVPPAEAEAT							339
Db	435	SDGRSCEDLNCECESS							449

RF	SEQUENCE FROM N.A.
PC	TTSUE-Skin:

FT	DOMAIN	1113	1154	EGF-LIKE 17, CALCIUM-BINDING.
CT	DOMAIN	1155	1196	EGF-LIKE 18, CALCIUM-BINDING.

```
FT DOMAIN 1197 1237 EGF-LIKE 19, CALCITON-BINDING.
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCITON-BINDING.
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCITON-BINDING.
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCITON-BINDING.
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCITON-BINDING.
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCITON-BINDING.
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCITON-BINDING.
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCITON-BINDING.
FT REPEAT 1528 1605 TGFBP 4.
FT REPEAT 1606 1647 EGF-LIKE 27, CALCITON-BINDING.
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCITON-BINDING.
FT REPEAT 1689 1765 TGFBP 5.
FT REPEAT 1766 1807 EGF-LIKE 29, CALCITON-BINDING.
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCITON-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCITON-BINDING.
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCITON-BINDING.
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCITON-BINDING.
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCITON-BINDING.
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCITON-BINDING.
FT REPEAT 2055 2126 TGFBP 6.
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCITON-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCITON-BINDING.
FT DOMAIN 2206 2246 EGF-LIKE 38, CALCITON-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCITON-BINDING.
FT REPEAT 2291 2332 EGF-LIKE 40, CALCITON-BINDING.
FT REPEAT 2333 2401 TGFBP 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCITON-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCITON-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCITON-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCITON-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCITON-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCITON-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCITON-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 85 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 776 790 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
```

```
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1501 BY SIMILARITY.
FT DISULFID 1497 1512 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
```

Query Match 13.5%; Score 270; DB 1; Length 2871;
Best Local Similarity 24.8%; Pred. No. 1.8e-12;
Matches 107; Conservative 33; Mismatches 128; Indels 164; Gaps 25;

```
QY 19 PPAEPAKKPPPC-----HRCGLVDR-----NCGMDTAKKNGGGGTA 59
Db 1141 PPGHOLANISACIDINECELSAHLCPHRCVNLIGKYOCACNPGYHSTPDLRF----- 1194
QY 60 WEEKTLKYESESEIRLLEILEGLCES-----SDPECN-----OMLEAQEHLLEAMWLQ 108
Db 1195 -----CVDIDECISIMNGGCEFTPCTNSEGSEYSCQGFALMPQR----- 1234
QY 109 KSEYPDLEFMEFCVTKLYC-----CS-PTYGPDCLACOG-----GSORPCSGNGHC----- 154
Db 1235 --SCDIDDE--CEDPNICDGGQCTNIPGEY--RCLCYDGFMASEDMKTCVDVNECDLNP 1288
QY 155 ----SGD-GSNOGDSCHNGYOGPL-----CTDCMDGYFSSLRNETHSICTACDESKT 205
Db 1289 NICLSGTCEMKRGSFICHDGMYSGKRGKTCCTIDNECEIGHANHCDRHAIVCTMTAGSPK- 1347
QY 206 CSGLTNRDCCGCEGVWVDEGACVVDDECALEPPPCSAOFCKNANGSYTC----- 256
Db 1348 CS-----CSPEWIDGDKICTDLDSCSNGTHMCSOHADCKNTMGSTRLCKRGTYGD 1398
QY 257 -----BEC-----DSSCVGCTG-----EGP----- 271
Db 1399 GFTCTDDECESENILNLCNGGCLNAPGGYRCECDMGFPASDGRKCEDIDECSLPNICVF 1458
QY 272 GNCK-----ECISGTA--REHQCADVDECSLAETCYRKKNENCNTNGSYVCVCPD 321
Db 1459 GTCHNLPGLRCECEIGEILDRSGGNCYDVNEC-LDPTTCTI--SGNCVNTBGSTYCDCP 1515
QY 322 GFE--ETEDACV 331
```

DB 1516 DFEINPTFVGCV 1527

RESULT 4

FBNI_HUMAN STANDARD; PRT; 2871 AA.

AC P35355;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Fibrillin 1 precursor.

GN FBNI OR FBNI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

[1]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE=93372860; PubMed=8364578;

RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B., Pangillan T., Bonadio J.;

RT "Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome.";

RL Hum. Mol. Genet. 2:961-968(1993).

[2]

RP SEQUENCE OF 1-932 FROM N.A.

RC TISSUE=Placenta, and Fibroblast;

RX MEDLINE=94010947; PubMed=7691719;

RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;

RT "Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5' end.";

RL Genomics 17:476-484(1993).

[3]

RP SEQUENCE OF 899-2871 FROM N.A.

RX MEDLINE=91304368; PubMed=1852207;

RA Maslen C.L., Corson G.M., Maddox B.K., Gnanville R.W., Sakai L.Y.;

RT "Partial sequence of a candidate gene for the Marfan syndrome.";

RL Nature 352:334-337(1991).

[4]

RP SEQUENCE OF 813-1313 FROM N.A.

RX MEDLINE=91304367; PubMed=1852206;

RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M., Tsipouras P., Ramirez F., Hollister D.W.;

RT "Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes.";

RL Nature 352:330-334(1991).

[5]

RP CHARACTERIZATION.

RX MEDLINE=91317849; PubMed=1860873;

RA Sakai L.Y., Keene D.R., Gnanville R.W., Bachinger H.P.;

RT "Purification and partial characterization of fibrillin, a cysteine-rich structural component of connective tissue microfibrils.";

RL J. Biol. Chem. 266:14763-14770(1991).

[6]

RP STRUCTURE BY NMR OF 2054-2125.

RX MEDLINE=98031893; PubMed=9362480;

RA Yuan X., Downing A.K., Knott V., Handford P.A.;

RT "Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils.";

RL EMBO J. 16:659-666(1997).

[7]

RP STRUCTURE BY NMR OF 2124-2205.

RX MEDLINE=96144829; PubMed=8568869;

RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;

RT "Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1.";

RL J. Mol. Biol. 255:22-27(1996).

[8]

RP STRUCTURE BY NMR OF 2124-2205.

RX MEDLINE=96222301; PubMed=8653794;

RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D., Handford P.A.;

RT "Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders.";

RL Cell 85:597-605(1996).

[9]

RP REVIEW ON MFS VARIANTS.

RX MEDLINE=96174615; PubMed=8594563;

RA Colliod G., Beroud C., Soussi T., Junien C., Boileau C.;

RT "Software and database for the analysis of mutations in the human FBNI gene.";

RL Nucleic Acids Res. 24:137-141(1996).

[10]

RP REVIEW ON MFS VARIANTS.

RX MEDLINE=97169383; PubMed=9016526;

RA Colliod-Beroud G., Beroud C., Agés L., Black C., Boxer M., Brock D.J., Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L., Richards R.I., Wang W., Junien C., Boileau C.;

RT "Marfan Database (second edition): software and database for the analysis of mutations in the human FBNI gene.";

RL Nucleic Acids Res. 25:147-150(1997).

[11]

RP REVIEW ON VARIANTS.

RX MEDLINE=98062175; PubMed=9401003;

RA Hayward C., Brock D.J.H.;

RT "Fibrillin-1 mutations in Marfan syndrome and other type-1 fibrillinopathies.";

RL Hum. Mutat. 10:415-423(1997).

[12]

RP VARIANT MFS PRO-1137.

RX MEDLINE=91304569; PubMed=1852208;

RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y., Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J., Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;

RT "Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene.";

RL Nature 352:337-339(1991).

[13]

RP VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.

RX MEDLINE=93250834; PubMed=1301946;

RA Dietz H.C., Saraiya J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;

RT "Clustering of fibrillin (FBNI) missense mutations in Marfan syndrome patients at cysteine residues in EGF-like domains.";

RL Hum. Mutat. 1:366-374(1992).

[14]

RP VARIANT MFS SER-2307.

RX MEDLINE=92235290; PubMed=1569206;

RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr., Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A.;

RT "Marfan phenotype variability in a family segregating a missense mutation in the epidermal growth factor-like motif of the fibrillin gene.";

RL J. Clin. Invest. 89:1674-1680(1992).

[15]

RP VARIANTS MFS ILE-548 AND ALA-723.

RX MEDLINE=94010946; PubMed=8406497;

RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C., Pyeritz R.E., Francomano C.A.;

RT "Four novel FBNI mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome.";

RL Genomics 17:468-475(1993).

[16]

RP VARIANTS MFS SER-2144.

RX MEDLINE=93278402; PubMed=8504310;

RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;

RT "A novel fibrillin mutation in the Marfan syndrome which could disrupt calcium binding of the epidermal growth factor-like module.";

RL Hum. Mol. Genet. 2:475-477(1993).

[17]

RP VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.

RX MEDLINE=94108431; PubMed=8281141;

RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=96293359; PubMed=8698613;
 RA Constam D.B., Calton M., Robertson E.J.;
 RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone
 morphogenetic proteins at distinct sites during embryogenesis."; J.
 RL Cell Biol. 134:181-191(1996).
 RN [6]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=97436919; PubMed=9291583;
 RA Rancourt S.L., Rancourt D.E.;
 RT "Marine subunitin-like proteinase SPC6 is expressed during embryonic
 RT implantation, somitogenesis, and skeletal formation."; Dev.
 RL Genet. 21:75-81(1997).
 CC -1- FUNCTION: TO REPRESENT A WIDESPREAD ENDOPROTEASE ACTIVITY
 CC WITHIN THE CONSTRUCTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE
 CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
 CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
 CC OF GROWTH FACTORS.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-1'-ZAA BONDS, WHERE XAA
 CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -1- SUBCELLULAR LOCATION: PCS5A IS SECRETED THROUGH THE REGULATED
 CC SECRETORY PATHWAY. PCS5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 CC EARLY ENDOSOMES.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PCS5B/LONG (SHOWN HERE)
 CC AND PCS5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PCS5A IS EXPRESSED IN MOST TISSUES BUT IS MOST
 CC ABUNDANT IN THE INTESTINE AND ADRENALS. PCS5B IS EXPRESSED IN THE
 CC INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.
 CC -1- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,
 CC EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER,
 CC BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT
 CC E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT
 CC E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND
 CC NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK
 CC SAC FOLLOWED BY A CONFINATION TO DERMATOME COMPARTMENT. BETWEEN
 CC E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL
 CC CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED
 CC TO THE CONDENSING MESenchyme SURROUNDING THE CARPALS. AT THIS
 CC STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL
 CC CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,
 CC ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILI.
 CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF
 CC ISOFORM B OCCUR AT E12.5.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM.
 CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
 CC SORTING INFORMATION. AC 1 DIRECTS TEN LOCALIZATION AND INTERACTS
 CC WITH THE TGN SORTING PROTEIN PACS-1.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL: D17583; BAA04507.1; -;
 DR EMBL: D12619; BAA02143.1; -;
 DR EMBL: L14932; AAA74636.1; -;
 DR PIR: JX0248; JX0248.
 DR PIR: A48225; A48225.
 DR HSSP: O99405; 1MPF.
 DR MEROPS: S08.076; -;
 DR MGD; MGI:97515; PcsK5.
 DR InterPro: IPR000561; EGF-like.

[illegible]


```

FT CARBOHYD 537 537 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 541 541 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARSPLIC 569 705 EPROKDTVPCITSCRPNDACVADPHYTHYISLPTER
FT EFTREKDTITFLRAVTPILPANQADITFDTEGNLRSFDLI
FT KRYEDGMTVGVNQVRPIVGFPAVLKLENNYVIGVSHR
FT NVVNHIFVESEWF -> RCARLPCHEOCPRPLRTIY
FT HLESPNIOYPAVAFMRGPPSSAVGDGMOALATIGNEGEFF
FT TTRKVSHTSGCVALTTRKPIPEPRDLITVKMDIVRHGTVSSFF
FT VALKIFVSNEL (IN ISOFOR D).
SQ SEQUENCE 705 AA: 78056 MW: FD3F06469A4BAE2B CRC64:

Query Match 12.6%: Score 253.5; DB 1; Length 705;
Best Local Similarity 27.1%: Pred. No. 8.3e-12;
Matches 80; Conservative 24; Mismatches 108; Indels 83; Gaps 14:

QY 115 LFEWFCVTKLKCCSPGTYGPDCLACOGS-QRP-----CSGNHGCSG 156
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 131 MISYQGLVFRACCVKARENSDFVQNGADLDQPAKIPDEEDODPYLNDRCGGGPKQ 190
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 DGSNQGSG-SCRCHMGYQ-----GPLCTD---CMGYTS-----SLRNETHICTACDESC 203
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 191 QCRDPTGVETICSCFVGYQLSDGVSCEBDINECTGSHNCRLGESCINTVGSFRCQRSSC 250
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 KTCGSLTN---RDGCECEV-----GWVLDE-GACVD 230
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 251 GTGEELTEDNNCKRIDCEFCIHNCPPDFICQNTLGSFRCPKIQCKSGFIQDALGNCTD 310
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 VDECAEPSPCSAQAQFCRNANGSYTCECDSSVCYGTGEGPNCKECISG-AREHG-QC 288
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 311 INECTLSAPCPVQOTCINTFGSYTCOK-----NVPNCGRGYHLNDEGTRC 356
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 289 ADVDECSLAETCYCRKNKNCNTGSGYVCVCPDG--FEETEDACVPPAEATGEG 341
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 357 VDVECAPPAEPC-GKGHHCLNSPSCFCECKAGFYFDGISRTGVINECQRYPG 410
: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
FBNI_MOUSE STANDARD: PRT: 2871 AA.
ID FBNI_MOUSE
AC Q61534; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Fibrillin 1 precursor.
GN FBNI OR FBNI-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; Pubmed=7829516;
RA Yin W., Germiller J., Sanguinetti C., Smiley E., Panglilian T.,
RA Parella L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene.";
RL J. Biol. Chem. 270:1798-1806(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: L29454; AA556840.1; -
DR EMBL: U22493; AA64217.1; -
DR HSSP: P35555; IAPJ.
DR MGD: MGI:95489; Fbn1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002212; TB.
DR Pfam: PF00008; EGF; 46.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF-Ca; 42.
DR SMART: SM00001; EGF-like; 4.
DR PROSITE: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00022; EGF-1; 2.
DR PROSITE: PS01186; EGF-2; 38.
DR PROSITE: PS01187; EGF-Ca; 45.
DR KMW: Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat: Signal; Multigene family.
FT SIGNAL 1 27
FT CHAIN 1 2871
FT DOMAIN 1 81 112
FT DOMAIN 1 115 146
FT DOMAIN 1 147 178
FT DOMAIN 1 246 287
FT DOMAIN 1 288 329
FT DOMAIN 1 330 401
FT DOMAIN 1 402 446
FT DOMAIN 1 449 489
FT DOMAIN 1 490 529
FT DOMAIN 1 530 571
FT DOMAIN 1 572 612
FT DOMAIN 1 613 653
FT DOMAIN 1 656 721
FT DOMAIN 1 723 764
FT DOMAIN 1 765 806
FT DOMAIN 1 807 846
FT DOMAIN 1 910 951
FT DOMAIN 1 952 1018
FT DOMAIN 1 1028 1069
FT DOMAIN 1 1070 1112
FT DOMAIN 1 1113 1154
FT DOMAIN 1 1155 1196
FT DOMAIN 1 1197 1237
FT DOMAIN 1 1238 1279
FT DOMAIN 1 1280 1321
FT DOMAIN 1 1322 1363
FT DOMAIN 1 1363 1403
FT DOMAIN 1 1404 1445
FT DOMAIN 1 1446 1486
FT DOMAIN 1 1487 1527
FT DOMAIN 1 1528 1568
FT DOMAIN 1 1569 1607
FT DOMAIN 1 1608 1647
FT DOMAIN 1 1648 1688
FT DOMAIN 1 1689 1728
FT DOMAIN 1 1766 1807
FT DOMAIN 1 1808 1848
FT DOMAIN 1 1849 1890
FT DOMAIN 1 1891 1929
FT DOMAIN 1 1930 1972
FT DOMAIN 1 1973 2012
FT DOMAIN 1 2013 2054
FT DOMAIN 1 2055 2121
FT DOMAIN 1 2127 2165
FT DOMAIN 1 2166 2205
FT DOMAIN 1 2206 2246

```

```

FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.
FT DOMAIN 2333 2400 TGFBR 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 790 795 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.

```

```

FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.

Query Match 12.6% Score 252.5; DB 1; Length 2871;
Best Local Similarity 25.1%; Pred. No. 3.6e-11;
Matches 102; Conservative 31; Mismatches 124; Indels 149; Gaps 24;

QY 30 PCHRCGLVDKF-----NOGMVDTAKRFGGNTAMEEKLTSKESSEIRLLEILGLCES 85
DB 1167 PHRCVNLGKYGCACNPGVHPHDLRF-----CYDIDECSIMNGCET 1210
QY 86 -----SDRECN-----QMLEAOEHLLEAMWLQKSEYDLEFEMFCVTKLYC-----CS- 129
DB 1211 FCTNSDGSYECSCQPGFALMPDR-----SCRDIDQ--CEDNPNICDGGQCTN 1256
QY 130 -PETYGPDLACQG---GSQRPCSGNGHC-----SGD-GSRQDGSRCRMGYOGP 175
DB 1257 IPGEY--RCLCYDGFMASEDMKTCVYVNECDLNPATCLSGTCENRKGSTICDCMDGYSK 1314
QY 176 L---CTDCMDGYFSSLRNETHSICPACDESCATCGSLNRDQGECEGVWLDGACVDV 231
DB 1315 KGKTGCTDINECEIGHANGCRHAYCTNTAGSFK-CS-----CSPGMIGGKICTDL 1364
QY 232 DECAAEPPPCSAAPFCRKNANGSYTC-----EBC-----DSSVYCGT 267
DB 1365 DECSNGTHMCSOHADCKNTMGSYRCLCKDGYGDTCTDLDCEBSNMLNLCGGQCLNAP 1424
QY 268 G-----GEP-----GNCK-----ECISGYA--REHQ 287
DB 1425 GGYRCECDMGFVPSADGKACEDIDECSLPNICVFGCHNLPGLFRCECEIGYELDSGCGN 1484
QY 288 CADVDECSLAETQVRKNENCYTPGSYVCVCPDGE--ETEDACV 331
DB 1485 CTDVNEC-LDPTFCI--SGNCVCTPGSYTCDSCPDELNPRTVGCY 1527

RESULT 9
ID FBN2.MOUSE STANDARD; PRT; 2907 AA.
AC 061555; 063957;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Fibrillin 2 precursor.
GN FBN2 OR FBN-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95263670; PubMed=7744963;
RT Zhang H., Hu W., Ramirez F.;
of extracellular expression of fibrillin genes suggests heterogeneity
of extracellular microfibrils."

```

FT	DOMAIN	1442	1483	EGF-LIKE 24,	CALCIUM-BINDING
FT	DOMAIN	1484	1524	EGF-LIKE 25,	CALCIUM-BINDING
FT	DOMAIN	1525	1565	EGF-LIKE 26,	CALCIUM-BINDING
FT	REPEAT	1566	1642	TEGBP 4,	
FT	DOMAIN	1643	1684	EGF-LIKE 27,	CALCIUM-BINDING
FT	DOMAIN	1685	1726	EGF-LIKE 28,	CALCIUM-BINDING
FT	REPEAT	1727	1800	TEGBP 5,	
FT	DOMAIN	1801	1842	EGF-LIKE 29,	CALCIUM-BINDING
FT	DOMAIN	1843	1884	EGF-LIKE 30,	CALCIUM-BINDING
FT	DOMAIN	1885	1926	EGF-LIKE 31,	CALCIUM-BINDING
FT	DOMAIN	1927	1965	EGF-LIKE 32,	CALCIUM-BINDING
FT	DOMAIN	1966	2008	EGF-LIKE 33,	CALCIUM-BINDING
FT	DOMAIN	2009	2048	EGF-LIKE 34,	CALCIUM-BINDING
FT	DOMAIN	2049	2090	EGF-LIKE 35,	CALCIUM-BINDING
FT	REPEAT	2091	2163	TEGBP 6,	
FT	DOMAIN	2164	2205	EGF-LIKE 36,	CALCIUM-BINDING
FT	DOMAIN	2206	2245	EGF-LIKE 37,	CALCIUM-BINDING
FT	DOMAIN	2246	2286	EGF-LIKE 38,	CALCIUM-BINDING
FT	DOMAIN	2287	2330	EGF-LIKE 39,	CALCIUM-BINDING
FT	DOMAIN	2331	2372	EGF-LIKE 40,	CALCIUM-BINDING
FT	REPEAT	2373	2441	TEGBP 7,	
FT	DOMAIN	2442	2483	EGF-LIKE 41,	CALCIUM-BINDING
FT	DOMAIN	2484	2524	EGF-LIKE 42,	CALCIUM-BINDING
FT	DOMAIN	2525	2563	EGF-LIKE 43,	CALCIUM-BINDING
FT	DOMAIN	2564	2606	EGF-LIKE 44,	CALCIUM-BINDING
FT	DOMAIN	2607	2647	EGF-LIKE 45,	CALCIUM-BINDING
FT	DOMAIN	2647	2687	EGF-LIKE 46,	CALCIUM-BINDING
FT	DOMAIN	2688	2727	EGF-LIKE 47,	CALCIUM-BINDING
FT	DOMAIN	2728	1224	BY SIMILARITY.	
FT	DISULFID	115	127	BY SIMILARITY.	
FT	DISULFID	119	130	BY SIMILARITY.	
FT	DISULFID	132	141	BY SIMILARITY.	
FT	DISULFID	149	159	BY SIMILARITY.	
FT	DISULFID	153	164	BY SIMILARITY.	
FT	DISULFID	166	175	BY SIMILARITY.	
FT	DISULFID	180	190	BY SIMILARITY.	
FT	DISULFID	184	196	BY SIMILARITY.	
FT	DISULFID	198	207	BY SIMILARITY.	
FT	DISULFID	280	292	BY SIMILARITY.	
FT	DISULFID	287	301	BY SIMILARITY.	
FT	DISULFID	303	316	BY SIMILARITY.	
FT	DISULFID	322	334	BY SIMILARITY.	
FT	DISULFID	329	343	BY SIMILARITY.	
FT	DISULFID	345	358	BY SIMILARITY.	
FT	DISULFID	491	503	BY SIMILARITY.	
FT	DISULFID	498	512	BY SIMILARITY.	
FT	DISULFID	514	526	BY SIMILARITY.	
FT	DISULFID	532	542	BY SIMILARITY.	
FT	DISULFID	537	551	BY SIMILARITY.	
FT	DISULFID	553	566	BY SIMILARITY.	
FT	DISULFID	572	584	BY SIMILARITY.	
FT	DISULFID	579	593	BY SIMILARITY.	
FT	DISULFID	595	608	BY SIMILARITY.	
FT	DISULFID	614	625	BY SIMILARITY.	
FT	DISULFID	620	634	BY SIMILARITY.	
FT	DISULFID	636	649	BY SIMILARITY.	
FT	DISULFID	655	666	BY SIMILARITY.	
FT	DISULFID	661	675	BY SIMILARITY.	
FT	DISULFID	677	690	BY SIMILARITY.	
FT	DISULFID	765	777	BY SIMILARITY.	
FT	DISULFID	772	786	BY SIMILARITY.	
FT	DISULFID	788	801	BY SIMILARITY.	
FT	DISULFID	807	819	BY SIMILARITY.	
FT	DISULFID	814	828	BY SIMILARITY.	
FT	DISULFID	830	843	BY SIMILARITY.	
FT	DISULFID	849	859	BY SIMILARITY.	
FT	DISULFID	854	868	BY SIMILARITY.	
FT	DISULFID	870	883	BY SIMILARITY.	
FT	DISULFID	952	964	BY SIMILARITY.	
FT	DISULFID	959	973	BY SIMILARITY.	
FT	DISULFID	975	988	BY SIMILARITY.	
FT	DISULFID	1070	1082	BY SIMILARITY.	
FT	DISULFID	1077	1091	BY SIMILARITY.	
FT	DISULFID	1093	1106	BY SIMILARITY.	

```

FT DISULFID 1112 1124 BY SIMILARITY.
FT DISULFID 1119 1133 BY SIMILARITY.
FT DISULFID 1135 1149 BY SIMILARITY.
FT DISULFID 1155 1167 BY SIMILARITY.
FT DISULFID 1162 1176 BY SIMILARITY.
FT DISULFID 1178 1191 BY SIMILARITY.
FT DISULFID 1197 1209 BY SIMILARITY.
FT DISULFID 1204 1218 BY SIMILARITY.
FT DISULFID 1220 1233 BY SIMILARITY.
FT DISULFID 1239 1250 BY SIMILARITY.
FT DISULFID 1246 1259 BY SIMILARITY.
FT DISULFID 1261 1274 BY SIMILARITY.
FT DISULFID 1280 1292 BY SIMILARITY.
FT DISULFID 1303 1316 BY SIMILARITY.
FT DISULFID 1316 1334 BY SIMILARITY.
FT DISULFID 1322 1343 BY SIMILARITY.
FT DISULFID 1332 1358 BY SIMILARITY.
FT DISULFID 1345 1358 BY SIMILARITY.
FT DISULFID 1364 1377 BY SIMILARITY.
FT DISULFID 1371 1386 BY SIMILARITY.
FT DISULFID 1388 1399 BY SIMILARITY.
FT DISULFID 1405 1418 BY SIMILARITY.
FT DISULFID 1412 1427 BY SIMILARITY.
FT DISULFID 1429 1440 BY SIMILARITY.
FT DISULFID 1446 1458 BY SIMILARITY.
FT DISULFID 1453 1467 BY SIMILARITY.
FT DISULFID 1469 1482 BY SIMILARITY.
FT DISULFID 1488 1499 BY SIMILARITY.
FT DISULFID 1494 1508 BY SIMILARITY.
FT DISULFID 1510 1523 BY SIMILARITY.
FT DISULFID 1529 1540 BY SIMILARITY.
FT DISULFID 1535 1549 BY SIMILARITY.
FT DISULFID 1551 1564 BY SIMILARITY.
FT DISULFID 1647 1659 BY SIMILARITY.
FT DISULFID 1654 1668 BY SIMILARITY.
FT DISULFID 1670 1683 BY SIMILARITY.
FT DISULFID 1689 1701 BY SIMILARITY.

```

Query Match 12.6%; Score 252.5; DB 1; Length 2907;
 Best Local Similarity 24.8%; Pred. No. 3.7e-11;
 Matches 103; Conservative 30; Mismatches 123; Indels 159; Gaps 25;

```

OY 23 EAARPTPC-HRCGLVDFKNOGAVDTAKNFGGNTANEKTLKESSEIRLLEIEG 81
DB 2249 ECAONPLLCAFRCMTFGSYE-CTCPVGYG-LREDOCKMK-DLECAEG 2294
OY 82 L-CESSDEPCNOMLEAQBHEAMWLDKSEYRDLFEFVCYKTLKVCSPG-TYGPDL 138
DB 2295 LHDCEBRCMKCNL-----IGTFMCICPPGARRPDE 2327
OY 139 ACQGGSO--RP-CGNGHCSDGSGRQDSCRCMHGYQGPLC-TDCMDG----- 183
DB 2328 GGVDENECHTKPGICE-NRCV--NIISYRCECEBGRQSSSGECLDNNQGLCFAY 2383
OY 184 -----YFSLAN-ETHSICFACD----- 200
DB 2384 LQTMQMASSSRNLYTKSPC-CCDGRGNGHOCLELPLRGTAQYKKICPHGPGYATDGRD 2442
OY 201 -ESCTGCSGL-TNRDCE-----CEVGVLDGE--ACVVDCAEAPPCSAQAQ 248
DB 2443 IDECKVMSPLCTNGOCVNTMGSPRCFCCKAGYIMDISGTACVDDECSQSPKPCNF--ICK 2500
OY 249 NANGSYTC-----EECDSS-----CVGCTGRGPNCKECLISGYAR 283
DB 2501 NFKGSYOCSCPRGYVLDQEDGKTKDLDECQTKOHNOQGLCVMTLG--GFTCKCPRGFTQ 2557
OY 284 EHGQCADVDECSLAETKYRKNNENYNTPGSYVVCPRDGF-----ETDACC 330
DB 2558 HHTACIDNNEGCSOPSLCGAKG-ICONTGPSFSCCEQGRGFSIDASGLNCEVDDEC 2611

```

RESULT 10
 FBN2_HUMAN

```

ID FBN2_HUMAN STANDARD: PRT; 2911 AA.
AC P35556;
DR 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 2 precursor.
GN FBN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=94165150; PubMed=8120105;
RX Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguineti C.,
RA Bonadio J., Mecham R.P., Ramirez F.;
RT "Structure and expression of fibrillin-2, a novel microfibrillar
component preferentially located in elastic matrices."
RL J. Cell Biol. 124:855-863(1994).
RN [2]
RP MEDLINE=91304567; PubMed=1852206;
RX Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA Tsiipouras P., Ramirez F., Hollister D.;
RT "Linkage of Marfan syndrome and a phenotypically related disorder to
two different fibrillin genes."
RL Nature 352:330-334(1991).
RN [3]
RP VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.
RX MEDLINE=96083599; PubMed=7493032;
RA Putnam E.A., Zhang H., Ramirez F., Milewicz D.M.;
RT "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,
congenital contractual arachnoidactyly."
RL Nat. Genet. 11:456-458(1995).
RN [4]
RP VARIANTS CCA HIS-1114.
RX MEDLINE=98407789; PubMed=9737771;
RA Babcock D., Gasner C., Franke U., Maslan C.;
RT "A single mutation that results in an asp-to-his substitution and
partial exon skipping in a family with congenital contractual
arachnoidactyly."
RL Hum. Genet. 103:22-28(1998).
RN [5]
RP VARIANTS CCA PHE-1141 AND TRP-1252.
RX MEDLINE=20259236; PubMed=10797416;
RA Belleh S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,
RA Godfrey M.;
RT "Two novel fibrillin-2 mutations in congenital contractual
arachnoidactyly."
RL Am. J. Med. Genet. 92:7-12(2000).
CC -I- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
THAT BINDS CALCIUM. FIBRILLIN-2-CONTAINING MICROFIBRILS REGULATE
THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.
CC -I- DISEASE: DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTUAL
ARACHNOIDACTYLY (CCA) (ALSO KNOWN AS BEALS SYNDROME). CCA IS
PHENOTYPICALLY SIMILAR TO MARFAN SYNDROME, BUT DOES NOT EFFECT THE
AORTA AND THE EYES.
CC -I- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
EGF-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: U03272; AAA18950.1;
DR EMBL: X62009; -; NOT_ANNOTATED_CDS.
DR PIR: S17063; S17063.
DR PIR: S31101; S31101.

```

[illegible]

	DOMAIN	2694	2733	EGF-LIKE 47, CALCIUM-BINDING.
FT	DISULFID	115	124	BY SIMILARITY.
FT	DISULFID	119	130	BY SIMILARITY.
FT	DISULFID	132	141	BY SIMILARITY.
FT	DISULFID	149	159	BY SIMILARITY.
FT	DISULFID	153	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	180	190	BY SIMILARITY.
FT	DISULFID	184	195	BY SIMILARITY.
FT	DISULFID	197	206	BY SIMILARITY.
FT	DISULFID	279	291	BY SIMILARITY.
FT	DISULFID	286	300	BY SIMILARITY.
FT	DISULFID	302	315	BY SIMILARITY.
FT	DISULFID	321	333	BY SIMILARITY.
FT	DISULFID	328	342	BY SIMILARITY.
FT	DISULFID	344	357	BY SIMILARITY.
FT	DISULFID	497	509	BY SIMILARITY.
FT	DISULFID	504	518	BY SIMILARITY.
FT	DISULFID	520	532	BY SIMILARITY.
FT	DISULFID	538	548	BY SIMILARITY.
FT	DISULFID	543	557	BY SIMILARITY.
FT	DISULFID	559	572	BY SIMILARITY.
FT	DISULFID	578	590	BY SIMILARITY.
FT	DISULFID	585	599	BY SIMILARITY.
FT	DISULFID	601	614	BY SIMILARITY.
FT	DISULFID	620	631	BY SIMILARITY.
FT	DISULFID	626	640	BY SIMILARITY.
FT	DISULFID	642	655	BY SIMILARITY.
FT	DISULFID	661	672	BY SIMILARITY.
FT	DISULFID	667	681	BY SIMILARITY.
FT	DISULFID	683	696	BY SIMILARITY.
FT	DISULFID	771	783	BY SIMILARITY.
FT	DISULFID	778	792	BY SIMILARITY.
FT	DISULFID	794	807	BY SIMILARITY.
FT	DISULFID	813	825	BY SIMILARITY.
FT	DISULFID	820	834	BY SIMILARITY.
FT	DISULFID	836	849	BY SIMILARITY.
FT	DISULFID	855	865	BY SIMILARITY.
FT	DISULFID	860	874	BY SIMILARITY.
FT	DISULFID	876	889	BY SIMILARITY.
FT	DISULFID	958	970	BY SIMILARITY.
FT	DISULFID	965	979	BY SIMILARITY.
FT	DISULFID	961	994	BY SIMILARITY.
FT	DISULFID	1076	1088	BY SIMILARITY.
FT	DISULFID	1083	1097	BY SIMILARITY.
FT	DISULFID	1099	1112	BY SIMILARITY.
FT	DISULFID	1118	1130	BY SIMILARITY.
FT	DISULFID	1125	1139	BY SIMILARITY.
FT	DISULFID	1141	1155	BY SIMILARITY.
FT	DISULFID	1161	1173	BY SIMILARITY.
FT	DISULFID	1168	1182	BY SIMILARITY.
FT	DISULFID	1184	1197	BY SIMILARITY.
FT	DISULFID	1203	1215	BY SIMILARITY.
FT	DISULFID	1210	1224	BY SIMILARITY.

Query Match 12.4% Score 249.5; DB 1; Length 2911;
Best Local Similarity 21.1%; Pred. No. 6; 2e-11;
Matches 100; Conservative 53; Mismatches 125; Indels 197; Gaps 26

OY	33	RGRGVADRF-----NOGM-VDTAKN-----FGGG--NT--AMEKITSKYE	69
Dd	1087	KRRNIIGSKRCCKNSGFALDMEERNCTDIDECRISPDLCGSGCVNPFSFECEGEYE	1146
OY	70	SSEIRLLEILE-----GLCESSD--FEC-----NQMLEAQBHLAEAWMLQK	109
Dd	1147	SGFMKKMKNCIDIDGEKNPLLCRGGYCVNTEBSFOQDCPLGHBLSPSRDVCVDINCSLS	1206
OY	110	SEYPDLFEWFC-----VKTLLKCCSGGTG-PD-----CLACQGSGQRPCSN	151
Dd	1207	DN-----LCRNGKCVCNNIIGTYQCSCNPGYATPDRGCTDIDECMTMNGGCDTQCTWS	1259
OY	152	GHCSGDGSGROGDSGRCHMGY-----OGPICTD-----CM	181

Db 1260 -----EGSYECSCSEGYALMPDGRSCADIDECENNEDIDCGQCTINIPGEYRCLEY 1310
 QY 182 DGYFSLRNETHSICTACDESKCTC-----SGLTN-RDCEC 217
 Db 1311 DGFNMAKDKTCTIDVNECDLNSNICMGECEKNTKGSFICHCOLGYVKKCTCTCTDVEDE 1370
 QY 218 EV-----GWVLDGACVYDVECAAEPPCSAQAFCNKANGS 253
 Db 1371 EIGAHCNDMAHSLNIPGSKSCSRGAWIGNGIKICIDLECSNGTHQCSINACVMTPGS 1430
 QY 254 YFCEEDSSCVGTGEG-----PENCK-ECISGY--AREHG 286
 Db 1431 YRC-ACSE---GFTGDFTCSDVDECAENINLCEGQCLNVPAYRCCECMGFTTPASDSR 1486
 QY 287 QCADVDECSIAEKTVCYRKNECYNTPGYSVYCVCPDFEETE-----DACVPP 333
 Db 1487 SCQDIDECSE-ONICV--SQTGNLPGMFHCICIDDEGYELDRGNGCTDIDECADP 1538
 RESULT 11
 PKCS_BRACL STANDARD; PRT: 1696 AA.
 AC 09NJ15: 09NJ16: 09NJ14:
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
 GN (Proprotein convertase PC6-like) (apc6).
 OS Branchiostoma californiensis (California lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 RX NCB1_TaxID=738;
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
 RA MEDLINE=20175281; PubMed=10708868;
 RA Oliva A.A. Jr., Chan S.O., Steiner D.F.;
 RT Evolution of the prothormone convertases: Identification of a
 RT homologue of PC6 in the protochordate amphioxus."
 RL Biochim. Biophys. Acta 1477:338-348(2000).
 CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
 CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 CC OF CLEAVAGE AT THE RK(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
 CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
 CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF184615; AAF26300.1; -
 DR EMBL: AF184616; AAF26301.1; -
 DR EMBL: AF184617; AAF26302.1; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002884; P_domain.
 DR InterPro: IPR000209; Peptidase_S8.
 DR PRINTS: PR00723; SUBTILISIN.

DR ProDom; PD000717; P_domain; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00261; FU; 17.
 DR PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR Hydroxylase; Serine protease; Glycoprotein; Zymogen; Signal;
 KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
 KW Transmembrane.
 FT SIGNAL 1 25
 FT PROPEP 26 110
 FT CHAIN 111 1696
 FT DOMAIN 111 1618
 FT TRANSMEM 1619 1639
 FT DOMAIN 1640 1696
 FT DOMAIN 111 488
 FT DOMAIN 496 637
 FT DOMAIN 664 1649
 FT SITE 110 111
 FT ACT_SITE 192 192
 FT ACT_SITE 233 233
 FT ACT_SITE 407 407
 FT CARBOHYD 246 246
 FT CARBOHYD 529 529
 FT CARBOHYD 885 885
 FT VARSPLIC 1259 1323
 FT FT
 FT FT
 FT VARSPLIC 1344 1696
 FT SEQUENCE 1696 AA; 188410 MW; 281CBEL784257CBD CRC64;
 SQ

Query Match 12.4%; Score 248.5; DB 1; Length 1696;
 Best Local Similarity 19.9%; Pred No. 4.4e-11;
 Matches 78; Conservative 40; Mismatches 101; Indels 173; Gaps 16;

QY 89 EKNOMLE-NOEHLFAMWLQKSEYDLEFEMFCVYTLKVCSPGTYG-----P 135
 Db 938 QCNQCTIDYEGEDHF-----LYGTC-----HVTPPGLYGFTTQVYKACAP 980
 QY 136 DCLACGGGSGQRCS-----GNHCSGDSGRQDGSRR--CHMGYGP----- 175
 Db 981 GGIACDGPADNOCITLCEERAPLDGRCSQSDTEARFACBCHSCSEGPDIQDSCDEY 1040
 QY 176 -----LCTDCMGYFSS----- 187
 Db 1041 YLTEDTCVARTNCPSTFYDDDDRECRPHDNCCEACDGNNONCNCSCKGFKTPDGCST 1100
 QY 188 -----LRNETHSICTACDESKCTSGLTNRDCEGEVWVLDGAC-----VD 230
 Db 1101 GCPNRYKDDTJKCKPCDSSCFCTGSGPASFPCLSCADDPLHESCRSTCPAGFIGNNE 1160
 QY 231 VDECAAEPPCSAQAFCNKANGSYTCEDSSCVGTGSGPNKCECISGVYRHHGQCAD 290
 Db 1161 SHRCVSS--SCRDQDYYSSETG--RCEDCPYNCRC--DNDDCAECAPTYYVVGRCRP 1214
 QY 291 VDECAAEPPCSAQAFCNKANGSYTCEDSSCVGTGSGPNKCECISGVYRHHGQCAD 312
 Db 1215 EETCEDGEYQDRDRTAELSCRPCHQSCCTGSGPSDTPDCSCGDDTILDRGECTITSGP 1274
 QY 313 GSIV-----C-VCPDGEFETDAC 330
 Db 1275 GEYMDRREKKKCAKACHTCKECGDEYDDTCTAC 1306
 RESULT 12

PRS-RAT
 ID PRS-RAT STANDARD: PRT: 675 AA.
 AC P53813;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Vitamin K-dependent protein S precursor.
 GN PROS1 OR PROS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Rattus.
 NC NCBL_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Liver;
 RA MEDLINE-9533263; PubMed-7608128;
 RX Yasuda F., Haysht T., Tanitame K., Nishioka J., Suzuki K.:
 RT "Molecular cloning and functional characterization of rat plasma
 protein S.";
 RL J. Biochem. 117:374-383(1995).
 CC - FUNCTION: ANTICOAGULANT PLASMA PROTEIN. IT IS A COFACTOR TO
 CC ACTIVATED PROTEIN C IN THE DEGRADATION OF COAGULATION FACTORS VA
 CC AND VIIIA. IT HELP TO PREVENT COAGULATION AND STIMULATING
 CC FIBRINOLYSIS.
 CC - SUBCELLULAR LOCATION: Extracellular.
 CC - TISSUE SPECIFICITY: PLASMA.
 CC - SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 CC - SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S78744; AAC60704.1; -
 DR HSSP: P00740; ICFH.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002383; GLA_blood.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR000294; VltK_dep_GLA.
 DR Pfam: PF00008; EGF; 4.
 DR Pfam: PF00594; gla; 1.
 DR Pfam: PF00054; laminin_G; 1.
 DR PRINTS: PR00001; GLABLOOD.
 DR SMART: SM00179; EGF_CA; 3.
 DR SMART: SM00001; EGF_Like; 1.
 DR SMART: SM00069; GLA; 1.
 DR SMART: SM00282; LamG; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 4.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS00186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 3.
 DR PROSITE: PS00011; GLU CARBOXYLATION; 1.
 DR PROSITE: PS00025; LAM_G_DOMAIN; 1.
 KW Plasma; Gamma-carboxyglutamic acid; Calcium; Vitamin K; Zymogen;
 KW Blood coagulation; Hydroxylation; Glycoprotein; Signal; Repeat;
 KW EGF-like domain.
 FT SIGNAL 1 24
 FT PROPEP 25 41
 FT CHAIN 42 675
 FT DOMAIN 42 675
 FT DOMAIN 87 86
 FT DOMAIN 87 116
 FT DOMAIN 117 155
 FT DOMAIN 157 200
 FT DOMAIN 157 242
 FT DOMAIN 201 242
 FT DOMAIN 243 283
 FT DOMAIN 299 475
 FT DOMAIN 484 665
 BY SIMILARITY.
 BY SIMILARITY.
 VITAMIN K-DEPENDENT PROTEIN S.
 GLA.
 THROMBIN-SENSITIVE.
 EGF-LIKE 1.
 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 LAMININ G-LIKE 1.
 LAMININ G-LIKE 2.

FT	MOD_RES	47	47	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	48	48	(BY SIMILARITY).
FT	MOD_RES	55	55	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	55	55	(BY SIMILARITY).
FT	MOD_RES	57	57	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	57	57	(BY SIMILARITY).
FT	MOD_RES	60	60	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	60	60	(BY SIMILARITY).
FT	MOD_RES	61	61	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	61	61	(BY SIMILARITY).
FT	MOD_RES	66	66	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	66	66	(BY SIMILARITY).
FT	MOD_RES	67	67	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	67	67	(BY SIMILARITY).
FT	MOD_RES	70	70	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	70	70	(BY SIMILARITY).
FT	MOD_RES	73	73	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	73	73	(BY SIMILARITY).
FT	MOD_RES	77	77	GAMMA-CARBOXYGLUTAMIC ACID
FT	MOD_RES	77	77	(BY SIMILARITY).
FT	MOD_RES	136	136	HYDROXYLATION (BY SIMILARITY).
FT	DISULFID	121	134	BY SIMILARITY.
FT	DISULFID	126	143	BY SIMILARITY.
FT	DISULFID	145	154	BY SIMILARITY.
FT	DISULFID	161	175	BY SIMILARITY.
FT	DISULFID	171	184	BY SIMILARITY.
FT	DISULFID	186	199	BY SIMILARITY.
FT	DISULFID	205	217	BY SIMILARITY.
FT	DISULFID	212	226	BY SIMILARITY.
FT	DISULFID	228	241	BY SIMILARITY.
FT	DISULFID	247	256	BY SIMILARITY.
FT	DISULFID	252	265	BY SIMILARITY.
FT	DISULFID	267	282	BY SIMILARITY.
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	509	509	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	675 AA;	74626 MW;	B4338F756B86075 CRC64;

Query Match 12.38; Score 247; DB 1; Length 675;
 Best Local Similarity 26.98; Pred. No. 2,4e-11;
 Matches 93; Conservative 34; Mismatches 113; Indels 106; Gaps 21;

QY	11	LLPLLLLPAPPAKKTPTCHRGGLVDRKNGQAVDTAKNFGGNTAMKEKTLKTES	70
DB	11	LLACLAIVLPNSFT-----NPLSKERASQVIVRRKR--ANTLLEETKKNLDR	56
QY	71	SEIRLEILLEGICSDPECNOMLEAOEHH-----LEAWML-----OLKSEYP	113
DB	57	-----ECIEELCNKE--EAREVEENNPEDYFPKYLGCIGARRVGAFAARSANAYP	108
QY	114	DLEFMCVKTLLKVCSPGTGPGDCLACQGGSGRPGSCNGHCS--GDGSRQDGGSCRGHMGY	172
DB	109	DLRS--CVNALPDQCDP-----MPCNEDGYLSCKDG--QGAFCTICKPGW	149
QY	173	QGRPLC-----TDCMGYFSSLRNETHSICTACDECKTCSGLTNDGCEVEGWL--DEG	226
DB	150	QGDGCGQPDINECKD-----PSNINGGCSQYTCDDTMPG--SYHC--SKIGFAMLTNNK	197
QY	227	ACVAVDECAAPPCSAAPFCCKNANGSVTECECCSCVCGTGECPGNCKECISGAREHG	286
DB	198	DKRYVDECSLKPSCVGR-VCKNIPGDFEC-----EPNRIYRDP	237
QY	287	--OCADVDECSLAEKTCVRKNENCNYPGSYVCV--PDGEETED	328
DB	238	SKSKCDVDECS--ENTCA--QLCVNYPGYSYCDCKKKGFKLAOD	278

RESULT 13
 FB12_MOUSE
 ID FB12_MOUSE STANDARD: PRT: 1221 AA.
 AC P37889; Q9WU12;

DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fibulin-2 precursor.
 GN FBLN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-35.
 RC TISSUE=Fibroblast;
 RX MEDLINE=94064787; PubMed=8245130;
 RA Pan T.-C., Sasaki T., Zhang R.-Z., Faessler R., Timpl R., Chu M.-L.;
 RT "Structure and expression of fibulin-2, a novel extracellular matrix
 RT protein with multiple EGF-like repeats and consensus motifs for
 RT calcium binding.";
 RL J. Cell Biol. 123:1269-1277(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99337686; PubMed=10406956;
 RA Graessle S., Sicot F.-X., Gotta S., Chu M.-L.;
 RT "Mouse fibulin-2 gene. Complete exon-intron organization and promoter
 RT characterization.";
 RL Eur. J. Biochem. 263:471-477(1999).
 CC -1- FUNCTION: ITS BINDING TO FIBRONECTIN AND SOME OTHER LIGANDS IS
 CC CALCIUM DEPENDENT.
 CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF FIBULIN-2 DUE TO THE
 CC PRESENCE OR ABSENCE OF A SINGLE EGF-LIKE (3) REPEAT MAY ARISE BY
 CC ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: COMPONENT OF BOTH BASEMENT MEMBRANES AND OTHER
 CC CONNECTIVE TISSUES.
 CC -1- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on ways
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X75285; CAA53040.1; -;
 DR EMBL; AF135253; AAD34456.1; -;
 DR EMBL; AF135253; AAD34456.1; JOINED
 DR EMBL; AF135240; AAD34456.1; JOINED
 DR EMBL; AF135241; AAD34456.1; JOINED
 DR EMBL; AF135242; AAD34456.1; JOINED
 DR EMBL; AF135243; AAD34456.1; JOINED
 DR EMBL; AF135244; AAD34456.1; JOINED
 DR EMBL; AF135245; AAD34456.1; JOINED
 DR EMBL; AF135246; AAD34456.1; JOINED
 DR EMBL; AF135247; AAD34456.1; JOINED
 DR EMBL; AF135248; AAD34456.1; JOINED
 DR EMBL; AF135249; AAD34456.1; JOINED
 DR EMBL; AF135250; AAD34456.1; JOINED
 DR EMBL; AF135251; AAD34456.1; JOINED
 DR EMBL; AF135252; AAD34456.1; JOINED
 DR PIR; A49457; A49457.
 DR HSSP; P07204; IFGD.
 DR MGD; MGI:95488; Fbln2.
 DR InterPro; IPR0000020; Anaphylatoxin.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR Pfam; PF00008; EGF_6.
 DR Pfam; PF00008; EGF_6.
 DR SMART; SMART; ANAT0; 3.
 DR SMART; SMART; EGF_CA; 9.
 DR SMART; SMART; EGF_like; 2.
 DR SMART; SMART; EGF_CA; 9.
 DR SMART; SMART; EGF_like; 2.

DR	PROSITE, PS00010; ASX_HYDROXYL, 5.
DR	PROSITE, PS01177; ANAPHYLATOXIN_1; 3.
DR	PROSITE, PS01178; ANAPHYLATOXIN_2; 3.
DR	PROSITE, PS00022; EGF_1; FALSE_NEG.
DR	PROSITE, PS01166; EGF_2; 5.
DR	PROSITE, PS01167; EGF_CA_10.
KW	Signal; Glycoprotein; Extracellular matrix; Plasma; EGF-like domain;
RK	Calcium-binding; Alternate splicing; Repeat.
FT	SIGNAL 1
FT	CHAIN 27
FT	DOMAIN 27
FT	DOMAIN 27
FT	DOMAIN 177
FT	DOMAIN 435
FT	DOMAIN 478
FT	DOMAIN 511
FT	DOMAIN 594
FT	DOMAIN 669
FT	DOMAIN 709
FT	DOMAIN 756
FT	DOMAIN 801
FT	DOMAIN 847
FT	DOMAIN 895
FT	DOMAIN 938
FT	DOMAIN 980
FT	DOMAIN 1019
FT	DOMAIN 1062
FT	DOMAIN 1111
FT	SIZE 421
FT	DISULFID 435
FT	DISULFID 436
FT	DISULFID 449
FT	DISULFID 479
FT	DISULFID 492
FT	DISULFID 511
FT	DISULFID 512
FT	DISULFID 525
FT	DISULFID 598
FT	DISULFID 606
FT	DISULFID 621
FT	DISULFID 623
FT	DISULFID 679
FT	DISULFID 694
FT	DISULFID 713
FT	DISULFID 720
FT	DISULFID 742
FT	DISULFID 805
FT	DISULFID 812
FT	DISULFID 833
FT	DISULFID 899
FT	DISULFID 906
FT	DISULFID 923
FT	DISULFID 942
FT	DISULFID 950
FT	DISULFID 965
FT	DISULFID 984
FT	DISULFID 989
FT	DISULFID 1004
FT	DISULFID 1023
FT	DISULFID 1031
FT	DISULFID 1046
FT	DISULFID 1066
FT	DISULFID 1073
FT	DISULFID 1093
FT	CARBOHYD 179
FT	CARBOHYD 497
FT	CARBOHYD 737
FT	CARBOHYD 1072
FT	VARSPLIC 709
FT	CONFLICT 140
FT	CONFLICT 348
FT	CONFLICT 507
DR	PROSITE, PS00010; ASX_HYDROXYL, 5.
DR	PROSITE, PS01177; ANAPHYLATOXIN_1; 3.
DR	PROSITE, PS01178; ANAPHYLATOXIN_2; 3.
DR	PROSITE, PS00022; EGF_1; FALSE_NEG.
DR	PROSITE, PS01166; EGF_2; 5.
DR	PROSITE, PS01167; EGF_CA_10.
KW	Signal; Glycoprotein; Extracellular matrix; Plasma; EGF-like domain;
RK	Calcium-binding; Alternate splicing; Repeat.
FT	SIGNAL 26
FT	CHAIN 1221
FT	DOMAIN 27
FT	DOMAIN 27
FT	DOMAIN 176
FT	DOMAIN 434
FT	DOMAIN 477
FT	DOMAIN 478
FT	DOMAIN 510
FT	DOMAIN 543
FT	DOMAIN 635
FT	DOMAIN 708
FT	DOMAIN 755
FT	DOMAIN 800
FT	DOMAIN 846
FT	DOMAIN 894
FT	DOMAIN 937
FT	DOMAIN 979
FT	DOMAIN 1018
FT	DOMAIN 1061
FT	DOMAIN 1106
FT	DOMAIN 1121
FT	SIZE 421
FT	DISULFID 435
FT	DISULFID 436
FT	DISULFID 449
FT	DISULFID 479
FT	DISULFID 492
FT	DISULFID 511
FT	DISULFID 512
FT	DISULFID 525
FT	DISULFID 598
FT	DISULFID 610
FT	DISULFID 619
FT	DISULFID 621
FT	DISULFID 623
FT	DISULFID 679
FT	DISULFID 694
FT	DISULFID 707
FT	DISULFID 713
FT	DISULFID 720
FT	DISULFID 735
FT	DISULFID 754
FT	DISULFID 805
FT	DISULFID 812
FT	DISULFID 827
FT	DISULFID 833
FT	DISULFID 845
FT	DISULFID 899
FT	DISULFID 906
FT	DISULFID 921
FT	DISULFID 936
FT	DISULFID 954
FT	DISULFID 963
FT	DISULFID 978
FT	DISULFID 984
FT	DISULFID 993
FT	DISULFID 1002
FT	DISULFID 1017
FT	DISULFID 1023
FT	DISULFID 1035
FT	DISULFID 1044
FT	DISULFID 1060
FT	DISULFID 1079
FT	DISULFID 1088
FT	DISULFID 1105
FT	CARBOHYD 179
FT	CARBOHYD 497
FT	CARBOHYD 737
FT	CARBOHYD 1072
FT	VARSPLIC 709
FT	CONFLICT 140
FT	CONFLICT 348
FT	CONFLICT 507
DR	PROSITE, PS00010; ASX_HYDROXYL, 5.
DR	PROSITE, PS01177; ANAPHYLATOXIN_1; 3.
DR	PROSITE, PS01178; ANAPHYLATOXIN_2; 3.
DR	PROSITE, PS00022; EGF_1; FALSE_NEG.
DR	PROSITE, PS01166; EGF_2; 5.
DR	PROSITE, PS01167; EGF_CA


```

FT DISULFID 634 645 .BY SIMILARITY.
FT DISULFID 640 654 .BY SIMILARITY.
FT DISULFID 657 669 .BY SIMILARITY.
FT DISULFID 675 686 .BY SIMILARITY.
FT DISULFID 681 695 .BY SIMILARITY.
FT DISULFID 698 709 .BY SIMILARITY.
FT DISULFID 715 726 .BY SIMILARITY.
FT DISULFID 721 735 .BY SIMILARITY.
FT DISULFID 737 750 .BY SIMILARITY.
FT DISULFID 756 767 .BY SIMILARITY.
FT DISULFID 762 776 .BY SIMILARITY.
FT DISULFID 778 791 .BY SIMILARITY.
FT DISULFID 797 808 .BY SIMILARITY.
FT DISULFID 803 817 .BY SIMILARITY.
FT DISULFID 819 832 .BY SIMILARITY.
FT DISULFID 838 850 .BY SIMILARITY.
FT DISULFID 845 859 .BY SIMILARITY.
FT DISULFID 861 873 .BY SIMILARITY.
FT DISULFID 879 891 .BY SIMILARITY.
FT DISULFID 885 900 .BY SIMILARITY.
FT DISULFID 902 915 .BY SIMILARITY.
FT DISULFID 921 933 .BY SIMILARITY.
FT DISULFID 927 942 .BY SIMILARITY.
FT DISULFID 944 957 .BY SIMILARITY.
FT DISULFID 963 975 .BY SIMILARITY.
FT DISULFID 970 984 .BY SIMILARITY.
FT DISULFID 986 1000 .BY SIMILARITY.
FT DISULFID 1101 1114 .BY SIMILARITY.
FT DISULFID 1109 1123 .BY SIMILARITY.
FT DISULFID 1125 1138 .BY SIMILARITY.
FT DISULFID 1144 1155 .BY SIMILARITY.
FT DISULFID 1150 1164 .BY SIMILARITY.
FT DISULFID 1166 1179 .BY SIMILARITY.
FT DISULFID 1298 1309 .BY SIMILARITY.
FT DISULFID 1304 1318 .BY SIMILARITY.
FT DISULFID 1320 1333 .BY SIMILARITY.
FT DISULFID 1339 1354 .BY SIMILARITY.
FT DISULFID 1349 1363 .BY SIMILARITY.
FT DISULFID 1365 1378 .BY SIMILARITY.
FT MOD_RES 647 647 .HYDROXYLATION.
FT MOD_RES 810 810 .HYDROXYLATION.
FT CARBOHYD 21 21 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 52 52 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 294 294 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 870 870 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 923 923 .N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1039 1039 .N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1394 AA; 152791 MW; DFCAB1A40B2C7D1 CRC64;

```

```

Query Match 12.08; Score 240.5; DB 1; Length 1394;
Best Local Similarity 27.18; Pred. No. 1.5e-10;
Matches 85; Conservative 22; Mismatches 96; Indels 111; Gaps 19;

```

```

QY 67 KYESSEIRLEILEGCESSDFECNOMLEAOEHLLEAMWLOLKEYSFDLFEMPEVKILKV 126
DB 655 EYCDSTGRMKTQ--RGKCEDID--ECLNLPSTCPDEO-----CVN----- 688
QY 127 CCSFGTYGPDLACOGGSGORPCSGNGHC-----SGDGSR--QGDGSGCRHMGY 172
DB 689 --SPGSGY--QCVPTGEGFR--GMNGOCLDVEDLEPRVCAANGCNSNLESGYSKCHGY 741
QY 173 QGPICTCMGYSRLNETHSICTACDE-----SKTCSGLTNRDCGCEGVGW 222
DB 742 -----TRTPDHKHCRIIDECQGNLCVNGQCKNTGSGFRCTCGQ---GYQ 783
QY 223 LD--EGACVDVDECAARPPPCSAOPCKNANGSYTCERDSSCVGCGEGGNGKKECISG 280
DB 784 LSAKDCEEDIDEC--QHRHLCAHGO--CRNTEGSGFQC--VCDP---GYRASGLGD----- 830
QY 281 YAREHGOCADVDECSLAETCTVRKNENCYNTPGSGYVCVCPDGF-----ETE 327

```

```

DB 831 -----HCEIDNECTLEDKSKVQOR--GDCINTAGSYDCTCPDGFOLDNKTQCODINEENHP 882
QY 328 DACYPAPAEATEEG 341
DB 883 GLCGPQEGECUNTEG 896

RESULT 16
ID PRIS_HUMAN STANDARD; PRT; 676 AA.
AC P07225; Q15518;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vitamin K-dependent protein S precursor.
GN PROS1 OR PROS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91084444; Pubmed=2148110;
RA Schmidt D.K., Tatro A.V., Phelps L.G., Tomczak J.A., Long G.L.;
RT "Organization of the human protein S genes.";
RL Biochemistry 29:7845-7852(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91084445; Pubmed=2148111;
RA Ploos van Amstel H.K., Reitsma P.H., der Logt C.P., Bertina R.M.;
RT "Intron-exon organization of the active human protein S gene PS alpha
and its pseudogene PS beta: duplication and silencing during primate
evolution.";
RL Biochemistry 29:7853-7861(1990).
RN [4]
RP SEQUENCE OF 1-42 FROM N.A.
RX MEDLINE=88005138; Pubmed=2820795;
RA Ploos van Amstel H.K., van der Zanden A.L., Reitsma P.H.,
RA Bertina R.M.;
RT "Human protein S cDNA encodes Phe-16 and Tyr 222 in consensus
sequences for the post-translational processing.";
RL FEBS Lett. 222:186-190(1987).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092407; Pubmed=3467362;
RA Hoskins J., Norman D.K., Beckmann R.J., Long G.L.;
RT "Cloning and characterization of human liver cDNA encoding a protein
S precursor.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:349-353(1987).
RN [6]
RP SEQUENCE OF 27-676 FROM N.A.
RX MEDLINE=86313649; Pubmed=2944113;
RA Lundwall A., Dackowski W., Cohen E., Shaffer M., Mahr A., Dahlback B.,
RA Stenflo J., Wydro R.;
RT "Isolation and sequence of the cDNA for human protein S, a regulator
of blood coagulation.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6716-6720(1986).
RN [7]
RP VARIANT HEERLEN.
RX MEDLINE=90335440; Pubmed=2143091;
RA Bertina R.M., Ploos van Amstel H.K., van Wijngaarden A.,
RA Coenen J., Leemhuis M.P., Deutz-Verlouw P.P., van der Linden I.K.,
RA Reitsma P.H.;
RT "Heerlen polymorphism of protein S, an immunologic polymorphism due
to dimorphism of residue 460.";
RL Blood 76:538-548(1990).
RN [8]

```


QY 71 SEIRLLELLEGLCSSDFECNOMLE--AOEHLLEAMWL-----OLKSEYP 113
 DB 57 -----ECIEELCKNKE--EAREFENDEPETYFYFKYLVCLRSPTGTLFTAROSTNAYP 108
 QY 114 DLFEMFCVKTLKVCSSPGTYGPDCLACGGSGRPGSGNGHCS-GDGSFQDGGSCRCCHNGY 172
 DB 109 DLRS-CVNALPDCCSP-----LPCNEDGYMSCKKG--KASFTCTCKPGW 149
 QY 173 QGPIC-----TDCMDGYFSLNETHSICTACDESKTCSGLTNDCCGCEYGVWL--DEG 226
 DB 150 QGEKCEPDINCKD-----PSNINNGSGSQCIDNTPG--SYHC-SCKNGFVWLNNK 197
 QY 227 ACVAVDECAAPPPCSAOFCKNANGSYTCEECSSCVGCTGEGPNCCKECISGYA--RE 284
 DB 198 DCKADVDESLKPSICGTA-VCKNIPGDEP-----ECPEGTRYMLK 237
 QY 285 HGGCADDGCSLAETKCYRNKENCYNPGSYVYC--PDGEETED-----ACVP 332
 DB 238 SKSCEDIDECSS--ENMCA---QLCVNYPGYCYCDGKKGFKLADQDKSCCEVVSCLP 290

RESULT 17
 PCK5_RAT
 ID PCK5_RAT STANDARD: PRT: 1877 AA.
 AC P41413; 062914;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
 DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)
 DE (Convertase PC5) (PC5) (PC6) (Fragments).
 GN PCSK5.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND TISSUE-SPECIFICITY.
 RC TISSUE-Adrenal gland;
 RX MEDLINE=93342056; PubMed=9341687;
 RA Lussan J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
 RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a
 RT candidate proprotein convertase expressed in endocrine and
 RT nonendocrine cells.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
 RL [2]
 RP SEQUENCE OF 1676-1877 FROM N.A. (LONG ISOFORM).
 RC TISSUE-Adrenal gland;
 RA De Bie I., Marcinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=20214819; PubMed=10749928;
 RA Xiang Y., Molloy S.S., Thomas L., Thomas G.;
 RT "The PC6b cytoplasmic domain contains two acidic clusters that direct
 RT sorting to distinct trans-Golgi network/endosomal compartments.";
 RL Mol. Biol. Cell 11:1257-1273(2000).
 RN [4]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=97166043; PubMed=9013936;
 RA Zheng M., Seidah N.G., Plante J.E.;
 RT "The developmental expression in the rat CNS and peripheral tissues of
 RT proteases PCS and PACFA mRNAs: comparison with other proprotein
 RT processing enzymes.";
 RL Dev. Biol. 181:268-283(1997).
 CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
 CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE
 CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
 CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
 CC OF GROWTH FACTORS.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPEPTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA

CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -1- SUBCELLULAR LOCATION: PCS5 IS SECRETED THROUGH THE REGULATED
 CC SECRETORY PATHWAY. PCS5 IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 CC A PERINUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 CC EARLY ENDOSOMES (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PCS5B/LONG (SHOWN HERE)
 CC AND PCS5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE INTESTINE, BRAIN, ADRENAL
 CC GLAND, ANTERIOR PITUITARY, THYROID, OVARIES, TESTIS AND LUNG.
 CC HIGHEST LEVELS ARE FOUND IN THE GUT, DUODENUM, JEJUNUM AND ILEUM.
 CC EXPRESSION IS HIGHER IN FEMALE THAN IN MALE REPRODUCTIVE ORGANS.
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT E9 IN HIGHLY RESTRICTED
 CC REGIONS OF THE NEURAL TUBE, IN CAUDAL MYOTOMES, AND AT THE
 CC MATERNO-EMBRYONIC JUNCTION OF THE UTERUS. AT E10, RESTRICTED
 CC EXPRESSION IS DETECTED IN THE OPTIC AND OTIC VESTICLES, THE ROOF OF
 CC MIDBRAIN, AND TRUNK MYOTOMES. BY MIDGESTATION (E13-E16),
 CC EXPRESSION IN THE DEVELOPING NERVOUS SYSTEM HAS EXPANDED TO
 CC MULTIPLE REGIONS INCLUDING HIPPOCAMPUS, THALAMUS, HYPOTHALAMUS,
 CC BRAIN STEM, AND SPINAL CORD. EXPRESSION IS ALSO DETECTED IN
 CC SEVERAL PERIPHERAL ORGAN SYSTEMS, INCLUDING GUT, LUNG, ADRENAL AND
 CC KIDNEY PRIMORDIA.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM.
 CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
 CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS
 CC WITH THE TGN SORTING PROTEIN PACS-1.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation/
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L14933; AAA99906.1; -;
 CC EMBL: U47014; AAA87888.1; -;
 CC PIR: B48225; B48225.
 CC HSSP: Q99405; IMPT.
 CC MEROPS: S08.076; -;
 CC InterPro: IPR002174; Eurin-like.
 CC InterPro: IPR002884; P-domain.
 CC InterPro: IPR000209; Peptidase_S8.
 CC Pfam: PF01483; P. 1.
 CC Pfam: PF00082; Peptidase_S8; 1.
 CC PRINTS: PR00723; SUBTILISIN.
 CC PRODOM: PD000717; P-domain; 1.
 CC SMART: SM00261; FU; 6.
 CC PROSITE: PS00136; SUBTILASE_ASP; 1.
 CC PROSITE: PS00137; SUBTILASE_HIS; 1.
 CC PROSITE: PS00138; SUBTILASE_SER; 1.
 CC K01: Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;
 CC Cleavage on pair of basic residues; Alternative splicing; Repeat;
 CC Transmembrane.
 CC K01: Signal.
 CC FT SIGNAL .1 34
 CC FT PROPEP 35 116
 CC FT CHAIN 117 1877
 CC FT DOMAIN 117 1768
 CC FT TRANSMEM 1769 1789
 CC FT DOMAIN 1790 1877
 CC FT DOMAIN 117 452
 CC FT DOMAIN 464 602
 CC FT DOMAIN 638 1753
 CC FT DOMAIN 1825 1844
 CC FT DOMAIN 1856 1877
 CC FT SITE 116 117
 CC FT SITE 521 523
 CC FT ACT_SITE 173 173
 CC BY SIMILARITY.
 CC BY SIMILARITY.
 CC PROTEIN CONVERTASE SUBTILISIN/KEXIN
 CC TYPE 5.
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC CATALYTIC.
 CC HOMO B.
 CC CYS-RICH MOTIF (CRM) REGION.
 CC AC 1.
 CC AC 2.
 CC CLEAVAGE (AUTO-) (BY SIMILARITY).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).

```

FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 854 854 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1710 1710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1732 1732 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 878 915 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX ->
FT VARSPLIC 915 915 ATEESAEFGFCMLYKKNKLCQRKVLQQLCCKTCTGQ
FT VARSPLIC 916 1877 (IN ISOFORM PCSA).
FT VARSPLIC 1877 AA; 207888 MW; 890955DC60534444 CRC64;
SQ SEQUENCE

Query Match 11.8%; Score 236; DB 1; Length 1877;
Best Local Similarity 28.6%; Pred. No. 2.5e-10;
Matches 66; Conservative 16; Mismatches 79; Indels 70; Gaps 12;

QY 125 KVC---CSPCTYGPCLACGSGSRPCSGNGHCSGDSRGDSCRCRMHGYPGLCTDCM 181
DB 670 RIVVSSCPPEHFHADKKRC-----KCAPN-----CESCFESHADQCLSK 710
QY 182 DGYFSSLRNETHS-----ICTACDESKCTCSGLJTRNDGCECEVGVWIDE 225
DB 711 YGYF--LNERTSCVACPEGYODIKKIKGKSCSEKCTCTGFHN--CTECKGGLSLQG 766
QY 226 GACVDVDECAEPPPCSAOFCNKNANGSYCECDSSCVCTGEGPQNCKECTSGAREH 285
DB 767 SRC-----SVTCEDQGF--SGHCCQPCRCATCAGAGADGCTNCTEGYMER 813
QY 286 GQCADYDECSIA-----EKTGVKNENY--NTPGSYVC--VCPDGF 323
DB 814 GRC--VQSCSVSYLDHSLGEGYKSKCRCDNSCLTNGPFRKNCSCPSQY 862

RESULT 18
ID PCSK5_HUMAN STANDARD; PRT; 913 AA.
AC 092824; Q13527;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)
DE (Convertase PC5) (PC6) (hPC6).
GN PCSK5 OR PC5 OR PC6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RX MEDLINE=96353880; PubMed=8755538;
RA Miranda L., Wolf J., Pichuanes S., Duke R., Franzusoff A.;
RT Isolation of the human PC6 gene encoding the putative host protease
RT for HIV-1 gp160 processing in CD4+ T lymphocytes."
RL Proc. Natl. Acad. Sci. U.S.A. 93:7695-7700(1996).
RN [2]
RP REVISIONS.
RA Franzusoff A., Miranda L., Wolf J., Pichuanes S., Lu Y., Duke R.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 15-913 FROM N.A.
RA Reudelhuber T.L.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR

```

```

CC PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-2AA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC -!- SUBCELLULAR LOCATION: Secreted (by similarity).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS SEEMS TO BE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN T-LIMPHOCYTES.
CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAOLCULAR CHAPERONE
CC ASSISTING IN THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
DR EMBL: U56387; AAC50643.2; -
DR EMBL: U49114; AA91807.1; -
DR HSP: Q99405; IMP1.
DR MIM: 600488; -
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002884; P_domain.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF01483; P: 1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR ProDom: PD000717; P_domain; 1.
DR SMART: SM00261; FU; 5.
DR PROSITE: PS00136; SUBTILASE ASP; 1.
DR PROSITE: PS00137; SUBTILASE HIS; 1.
DR PROSITE: PS00138; SUBTILASE SER; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Cleavage on pair of basic residues; Repeat.
FT SIGNAL 1 32
FT PROPEP 33 114
FT CHAIN 115 913
FT TYPE 5.
FT CATALYTIC.
FT DOMAIN 115 454
FT DOMAIN 462 600
FT SITE 636 913
FT SITE 114 115
FT SITE 519 521
FT ACT_SITE 171 171
FT ACT_SITE 212 212
FT ACT_SITE 212 212
FT ACT_SITE 386 386
FT CARBOHYD 225 225
FT CARBOHYD 381 381
FT CARBOHYD 665 665
FT CARBOHYD 752 752
FT CARBOHYD 802 802
FT CARBOHYD 852 852
FT CONFLICT 118 118
FT CONFLICT 121 121
FT CONFLICT 511 511
FT CONFLICT 601 601
SQ SEQUENCE 913 AA; 101775 MW; 21389264CAD7546C CRC64;

Query Match 11.8%; Score 236; DB 1; Length 913;
Best Local Similarity 25.8%; Pred. No. 2.1e-10;
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKFNGWVDTRAKNGGNTAMEEKTLSKYSSEIRLEILGICLSSDFPCNDMLEAQ 97
DB 612 VERFRRSVREDDPDYDGTEDYAGP-----CDPECSV-----GCDGPDPHCNDCLH-- 658
QY 98 EHLLEAWMLQKSEYDLEDFEYVKTLL-----KVC-----CSP-----GTYGPCCLAC 140

```

Db 659 -----YYLKLNN-----TRICVSSCPGHHYHAKKRCRCAPNCESCFSGHGDQMSC 707
 Qy 141 OGG-----SQRPCSGNGHSCGDSRG--DGSRC-----CHMOYGPCLCTDCMDGYESSLRN 190
 Db 708 KYGFLNEEINSCYTHCP-DGSIYDTKKNLCKRCKSEKCTCTEHNCTECDGL--SLQG 764
 Qy 191 ETHSI-----CTACDESKCTCSGLTNRDCECEVWVLDBGACVYDECAAP 238
 Db 765 SRCVSCDEGRYFNGDCQCHRCATCAGAGADGCINCTGTYEMEDGRVQ----- 816
 Qy 239 PCGAAPFCNA--NGSYTCEECSSCVGCTGEGPGCKEISYAEHGO-----CAD 290
 Db 817 -SCSIYFPHSSSENGSKCKCDISLTCNGPSPFNCTSCPSGYLLDLGCMGGAICKD 875
 Qy 291 VDECSLAEK---TCVRKENCYNTPGSYVC 317
 Db 876 ATEESNAEGGFCMLYKKNKNCQKRVLDQIC 905

RESULT 19
 PRS_MOUSE STANDARD: PRT: 675 AA.
 ID PRS_MOUSE
 AC Q08761; P43483;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Vitamin K-dependent protein S precursor.
 GN PROS1 OR PROS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94198297; PubMed=8148380;
 RA Chu M.D., Sun J., Bird P.I.;
 RT Cloning and sequencing of a cDNA encoding the murine vitamin K-
 RT dependent protein S.";
 RL Biochim. Biophys. Acta 1217:325-328(1994).
 RN [2]
 RP SEQUENCE OF 33-675 FROM N.A.
 RX MEDLINE=94302659; PubMed=8029814;
 RA Lu D., Schmidt D.K., Long G.L.;
 RT Structure of mouse protein S as determined by PCR amplification and
 RT DNA sequencing of cDNA.";
 RL Thromb. Res. Suppl. 74:135-142(1994).
 CC -!- FUNCTION: ANTICOAGULANT PLASMA PROTEIN. IT IS A COFACTOR TO
 CC ACTIVATED PROTEIN C IN THE DEGRADATION OF COAGULATION FACTORS VA
 CC AND VIIIA. IT HELP TO PREVENT COAGULATION AND STIMULATING
 CC FIBRINOLYSIS.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: PLASMA.
 CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z25469; CAA80961.1; -;
 DR EMBL: L27439; AAA40006.1; -;
 DR HSSP: P00740; ICFH.
 DR MGD: MGI:1095733; Prosl.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002383; GLA_blood.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR000294; VitK_dep_GLA.

DR Pfam: PF00008; EGF; 4.
 DR Pfam: PF00594; gla; 1.
 DR Pfam: PF00054; laminin_G; 1.
 DR PRINTS: PR00001; GLABLOOD.
 DR SMART: SM00179; EGF_CA; 3.
 DR SMART: SM00001; EGF_like; 1.
 DR SMART: SM00069; GLA; 1.
 DR SMART: SM00282; Lamg; 2.
 DR PROSITE: PS00101; ASX_HYDROXYL; 4.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS01187; EGF_CA; 3.
 DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
 DR PROSITE: PS00025; LAM_G_DOMAIN; 2.
 DR Plasma: Gamma-carboxylglutamic acid: Calcium; Vitamin K; Zymogen;
 KW Blood coagulation; Hydroxylation; Glycoprotein; Signal; Repeat;
 KW EGF-like domain.
 FT SIGNAL 1 24
 FT PROPEP 25 41
 FT CHAIN 42 675
 FT DOMAIN 42 86
 FT DOMAIN 87 116
 FT DOMAIN 117 135
 FT DOMAIN 157 200
 FT DOMAIN 201 242
 FT DOMAIN 243 283
 FT DOMAIN 299 475
 FT DOMAIN 484 665
 FT MOD_RES 47 47
 FT MOD_RES 48 48
 FT MOD_RES 55 55
 FT MOD_RES 57 57
 FT MOD_RES 60 60
 FT MOD_RES 61 61
 FT MOD_RES 66 66
 FT MOD_RES 67 67
 FT MOD_RES 70 70
 FT MOD_RES 73 73
 FT MOD_RES 77 77
 FT MOD_RES 136 136
 FT DISULEID 121 134
 FT DISULEID 126 143
 FT DISULEID 145 154
 FT DISULEID 161 175
 FT DISULEID 171 184
 FT DISULEID 186 199
 FT DISULEID 205 217
 FT DISULEID 212 226
 FT DISULEID 228 241
 FT DISULEID 247 256
 FT DISULEID 252 265
 FT DISULEID 267 282
 FT CARBOHYD 499 499
 FT CARBOHYD 509 509
 FT CONFLICT 493 493
 FT SEQUENCE 675 AA; 74934 MW; 79D51203E5AF31F CRC64;

Query Match 11.6%; Score 232; DB 1; Length 675;
 Best Local Similarity 26.3%; Pred. No. 3; de-10;
 Matches 91; Conservative 30; Mismatches 119; Indels 106; Gaps 20;

OY 11 LRLPLLLPAPPAEAKRPPCHRCGLVDKFNQGVNVDKAKNKGNGNTAMEKTLKSYES 70
 DB 11 LLAALALVIVPSET-----NFLSKERASVLYVKRR---ANTLFEEETMKGNLER 56
 OY 71 SEIRLLIELEGCESSDFECNOMLEQEEH-----LEKMWL-----QLKSEFP 113
 DB 57 -----ECIEELCNKE--BAREVEFNENPPTDYFPYKICLGLAFRGVSHARQSANAP 108
 OY 114 DLFEMFCVTLKVCSPGYGPD-CLACOGSQRPCSGNHCSGDROGDSRCRCHMY 172
 DB 109 DLRS--CVMAISDQCPICNEGYLACODG-----QAATFCCKPCW 149
 OY 173 QGPPLC---TDCMDGYFSSLRNETHSICFACDESCCTKSGFLNRDCEGEVGVLT--DEG 226
 DB 150 QGDRCCQYDNECKD-----PSNVNGGCSQICDNTPG--SYHC--SCKRGFAMLPNKK 197
 OY 227 ACYVDVDECAEPPPCSAAPCKRANSGTYCEEDSSCVGTGEGPNCCEJCSGYARENG 286
 DB 198 DCKDLDECALKPSVCGTA-VCKNIPGDFEC-----ECPDGYRYDPS 237
 OY 287 --QCADVDECSLAECTCVKRNKENCYTPGSYVC--PDGFEETD 328
 DB 238 SKSKRVDECS--ENMCA---QLCVAFPGSGYSCYCDKKGFKLAQD 278

RESULT 20
 NID2_HUMAN STANDARD: PRT: 1375 AA.
 AC 014112: 043710:
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nidogen-2 precursor (NID-2) (Osteonidogen).
 GN NID2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=98406162; PubMed=9733643;
 RX Kohfeldt E., Sasaki T., Goehring W., Timpl R.;
 RT "Nidogen-2: a new basement membrane protein with diverse binding
 RT properties.";
 RL J. Mol. Biol. 282:99-109(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=canine bone;
 RA Ohno I., Hashimoto J., Yakaoka K., Ochi T., Okubo K., Matsubara K.;
 RT "The cloning and characterization of a cDNA for the novel bone matrix
 RT protein: osteonidogen.";
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ohno I., Okubo K., Matsubara K.;
 RT "Human osteonidogen gene: Intron-exon junctions and chromosomal
 RT localization.";
 RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: CELL ADHESION GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED
 CC IN BASEMENT MEMBRANES. BINDS TO COLLAGENS I AND IV, TO PERLECAN
 CC AND TO LAMININ 1. DOES NOT BIND FIBRILIN. IT PROBABLY HAS A ROLE
 CC IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.
 CC -1- TISSUE SPECIFICITY: HEART, PLACENTA AND BONE. LESS IN PANCREAS,
 CC KIDNEY AND SKELETAL MUSCLE.
 CC -1- PTM: HIGHLY N- AND O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-I DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ223500; CAA11418.1; -;
 DR EMBL; D64425; BAA13087.1; -;
 DR EMBL; AB009779; BAA24112.1; JOINED.
 DR EMBL; AB009778; BAA24112.1; JOINED.
 DR EMBL; AB009779; BAA24112.1; JOINED.
 DR EMBL; AB009780; BAA24112.1; JOINED.
 DR EMBL; AB009781; BAA24112.1; JOINED.
 DR EMBL; AB009782; BAA24112.1; JOINED.
 DR EMBL; AB009783; BAA24112.1; JOINED.
 DR EMBL; AB009784; BAA24112.1; JOINED.
 DR EMBL; AB009785; BAA24112.1; JOINED.
 DR EMBL; AB009786; BAA24112.1; JOINED.
 DR EMBL; AB009787; BAA24112.1; JOINED.
 DR EMBL; AB009788; BAA24112.1; JOINED.
 DR EMBL; AB009789; BAA24112.1; JOINED.
 DR EMBL; AB009790; BAA24112.1; JOINED.
 DR EMBL; AB009791; BAA24112.1; JOINED.
 DR EMBL; AB009792; BAA24112.1; JOINED.
 DR EMBL; AB009793; BAA24112.1; JOINED.
 DR EMBL; AB009794; BAA24112.1; JOINED.
 DR EMBL; AB009795; BAA24112.1; JOINED.
 DR EMBL; AB009796; BAA24112.1; JOINED.
 DR EMBL; AB009797; BAA24112.1; JOINED.
 DR EMBL; AB009798; BAA24112.1; JOINED.
 DR HSP; P35555; IEMN.
 DR MIM; 605399; -;
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF-like.
 DR InterPro; IPR000033; Ldl_receptor_rep.
 DR InterPro; IPR003886; Nidogen_ext.
 DR InterPro; IPR00716; Thyroglobulin_1.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF00058; ldl_recept_b; 4.
 DR Pfam; PF00086; thyroglobulin_1; 2.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00001; EGF_like; 3.
 DR SMART; SM00135; LY; 4.
 DR SMART; SM00539; NIDO; 1.
 DR SMART; SM00211; TY; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS01187; EGF_CA; 2.
 KW Basement membrane; Extracellular matrix; Glycoprotein; Signal;
 KW Calcium-binding; Repeat; EGF-like domain; Cell adhesion.
 FT SIGNAL 1 30
 FT CHAIN 31 1375 NIDOGEN-2.
 FT DOMAIN 484 524 EGF-LIKE 1.
 FT DOMAIN 759 800 EGF-LIKE 2.
 FT DOMAIN 801 843 EGF-LIKE 3.
 FT DOMAIN 848 891 EGF-LIKE 4.
 FT DOMAIN 932 930 EGF-LIKE 5.
 FT DOMAIN 936 1006 CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1015 1085 THYROGLOBULIN TYPE 1.
 FT DOMAIN 1157 1200 THYROGLOBULIN TYPE 1.
 FT DOMAIN 1201 1243 LDL-RECEPTOR YWTD MOTIF 1.
 FT DOMAIN 1244 1288 LDL-RECEPTOR YWTD MOTIF 2.
 FT DOMAIN 1289 1328 LDL-RECEPTOR YWTD MOTIF 3.
 FT DOMAIN 1329 1373 LDL-RECEPTOR YWTD MOTIF 4.
 FT DOMAIN 1373 1373 LDL-RECEPTOR YWTD MOTIF 5.
 FT DISULFID 763 776 BY SIMILARITY.
 FT DISULFID 770 786 BY SIMILARITY.
 FT DISULFID 788 799 BY SIMILARITY.
 FT DISULFID 805 818 BY SIMILARITY.
 FT DISULFID 812 827 BY SIMILARITY.
 FT DISULFID 829 842 BY SIMILARITY.
 FT DISULFID 852 867 BY SIMILARITY.
 FT DISULFID 859 877 BY SIMILARITY.
 FT DISULFID 879 890 BY SIMILARITY.


```

Db 1266 TCSQCVSRITNCNSKGLFQNGECRTTCADGYSD-----RGICACAKYLSCHTCSGPRR 1320
Qy 212 RDGECEVGVNLDEGACVDVDECAEPPPCSAAPFCNNANSYCECECDSCVCTEGEP 271
Db 1321 MOCVQCPAGWQLAAGEC-----HPECPEGFY-----KSDFCQCHYCKTKCNDAGR 1367
Qy 272 GNCKECTISGAREHGOCAD-----VDECSLAERTCVKRNKNCYNTPGSYCVCPDGFEE 325
Db 1368 LACTSCPHMLDGLCLMCECLSSQYDTTSATCTCHDCRSRSCG-PGQFSC----- 1418
Qy 326 TEDACVPP 333
Db 1419 --KGCVPP 1424

RESULT 23
NTCL_MOUSE
ID NTCL_MOUSE STANDARD; PRT; 2531 AA.
AC 001705;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (NOTCH protein).
GN NOTCH1 OR NOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Girdley T.;
RA Copeland N.G., Girdley T.;
RT Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homolog of Drosophila Notch.
RL Genomics 15:259-264(1993).
RN 12
RP SEQUENCE OF 1551-2170 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Girdley T.;
RT Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development.
RL Development 115:737-744(1992).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z11886; CAAT7941.1; -.
DR HSSP: P00740; IEDM.
DR MGD: MG1:97363; Notch1.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF_1like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR001438; EGF_IT.
DR InterPro: IPR000800; Notch.

```

```

DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 35.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00010; EGFBL00D.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 3.
DR SMART: SM00179; EGF_CA; 23.
DR SMART: SM00001; EGF_Like; 11.
DR SMART: SM00004; NL; 2.
DR PROSITE: PSS0088; ANK_REPEAT; 2.
DR PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 22.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01186; EGF_2; 27.
DR PROSITE: PS01187; EGF_CA; 21.
DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 2531
FT DOMAIN 19 1725
FT TRANSSEM 1726 1746
FT DOMAIN 1747 2531
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 450
FT DOMAIN 452 488
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 867
FT DOMAIN 869 905
FT DOMAIN 907 943
FT DOMAIN 945 981
FT DOMAIN 983 1019
FT DOMAIN 1021 1057
FT DOMAIN 1059 1095
FT DOMAIN 1097 1143
FT DOMAIN 1145 1181
FT DOMAIN 1183 1219
FT DOMAIN 1221 1265
FT DOMAIN 1267 1305
FT DOMAIN 1307 1346
FT DOMAIN 1348 1384
FT DOMAIN 1387 1426
FT DOMAIN 1449 1480
FT REPEAT 1445 1480
FT REPEAT 1481 1522
FT REPEAT 1523 1562
FT REPEAT 1562 1597
FT REPEAT 1597 1633
FT REPEAT 1633 1669
FT REPEAT 1669 1705
FT REPEAT 1705 1741
FT REPEAT 1741 1777
FT REPEAT 1777 1813
FT REPEAT 1813 1849
FT REPEAT 1849 1885
FT REPEAT 1885 1921
FT REPEAT 1921 1957
FT REPEAT 1957 1993
FT REPEAT 1993 2029
FT REPEAT 2029 2065
FT REPEAT 2065 2101
FT REPEAT 2101 2137
FT REPEAT 2137 2173
FT REPEAT 2173 2209
FT REPEAT 2209 2245
FT REPEAT 2245 2281
FT REPEAT 2281 2317
FT REPEAT 2317 2353
FT REPEAT 2353 2389
FT REPEAT 2389 2425
FT REPEAT 2425 2461
FT REPEAT 2461 2497
FT REPEAT 2497 2531
FT DISULFID 31 46
FT DISULFID 63 74
FT DISULFID 87 98
FT DISULFID 89 98
FT DISULFID 106 117

```

```

POTENTIAL.
NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
EGF-LIKE 1.
EGF-LIKE 2.
EGF-LIKE 3.
EGF-LIKE 4.
EGF-LIKE 5.
EGF-LIKE 6.
EGF-LIKE 7.
EGF-LIKE 8.
EGF-LIKE 9.
EGF-LIKE 10.
EGF-LIKE 11.
EGF-LIKE 12.
EGF-LIKE 13.
EGF-LIKE 14.
EGF-LIKE 15.
EGF-LIKE 16.
EGF-LIKE 17.
EGF-LIKE 18.
EGF-LIKE 19.
EGF-LIKE 20.
EGF-LIKE 21.
EGF-LIKE 22.
EGF-LIKE 23.
EGF-LIKE 24.
EGF-LIKE 25.
EGF-LIKE 26.
EGF-LIKE 27.
EGF-LIKE 28.
EGF-LIKE 29.
EGF-LIKE 30.
EGF-LIKE 31.
EGF-LIKE 32.
EGF-LIKE 33.
EGF-LIKE 34.
EGF-LIKE 35.
EGF-LIKE 36.
EGF-LIKE 37.
EGF-LIKE 38.
EGF-LIKE 39.
EGF-LIKE 40.
EGF-LIKE 41.
EGF-LIKE 42.
EGF-LIKE 43.
EGF-LIKE 44.
EGF-LIKE 45.
EGF-LIKE 46.
EGF-LIKE 47.
EGF-LIKE 48.
EGF-LIKE 49.
EGF-LIKE 50.
EGF-LIKE 51.
EGF-LIKE 52.
EGF-LIKE 53.
EGF-LIKE 54.
EGF-LIKE 55.
EGF-LIKE 56.
EGF-LIKE 57.
EGF-LIKE 58.
EGF-LIKE 59.
EGF-LIKE 60.
EGF-LIKE 61.
EGF-LIKE 62.
EGF-LIKE 63.
EGF-LIKE 64.
EGF-LIKE 65.
EGF-LIKE 66.
EGF-LIKE 67.
EGF-LIKE 68.
EGF-LIKE 69.
EGF-LIKE 70.
EGF-LIKE 71.
EGF-LIKE 72.
EGF-LIKE 73.
EGF-LIKE 74.
EGF-LIKE 75.
EGF-LIKE 76.
EGF-LIKE 77.
EGF-LIKE 78.
EGF-LIKE 79.
EGF-LIKE 80.
EGF-LIKE 81.
EGF-LIKE 82.
EGF-LIKE 83.
EGF-LIKE 84.
EGF-LIKE 85.
EGF-LIKE 86.
EGF-LIKE 87.
EGF-LIKE 88.
EGF-LIKE 89.
EGF-LIKE 90.
EGF-LIKE 91.
EGF-LIKE 92.
EGF-LIKE 93.
EGF-LIKE 94.
EGF-LIKE 95.
EGF-LIKE 96.
EGF-LIKE 97.
EGF-LIKE 98.
EGF-LIKE 99.
EGF-LIKE 100.
EGF-LIKE 101.
EGF-LIKE 102.
EGF-LIKE 103.
EGF-LIKE 104.
EGF-LIKE 105.
EGF-LIKE 106.
EGF-LIKE 107.
EGF-LIKE 108.
EGF-LIKE 109.
EGF-LIKE 110.
EGF-LIKE 111.
EGF-LIKE 112.
EGF-LIKE 113.
EGF-LIKE 114.
EGF-LIKE 115.
EGF-LIKE 116.
EGF-LIKE 117.
EGF-LIKE 118.
EGF-LIKE 119.
EGF-LIKE 120.
EGF-LIKE 121.
EGF-LIKE 122.
EGF-LIKE 123.
EGF-LIKE 124.
EGF-LIKE 125.
EGF-LIKE 126.
EGF-LIKE 127.
EGF-LIKE 128.
EGF-LIKE 129.
EGF-LIKE 130.
EGF-LIKE 131.
EGF-LIKE 132.
EGF-LIKE 133.
EGF-LIKE 134.
EGF-LIKE 135.
EGF-LIKE 136.
EGF-LIKE 137.
EGF-LIKE 138.
EGF-LIKE 139.
EGF-LIKE 140.
EGF-LIKE 141.
EGF-LIKE 142.
EGF-LIKE 143.
EGF-LIKE 144.
EGF-LIKE 145.
EGF-LIKE 146.
EGF-LIKE 147.
EGF-LIKE 148.
EGF-LIKE 149.
EGF-LIKE 150.
EGF-LIKE 151.
EGF-LIKE 152.
EGF-LIKE 153.
EGF-LIKE 154.
EGF-LIKE 155.
EGF-LIKE 156.
EGF-LIKE 157.
EGF-LIKE 158.
EGF-LIKE 159.
EGF-LIKE 160.
EGF-LIKE 161.
EGF-LIKE 162.
EGF-LIKE 163.
EGF-LIKE 164.
EGF-LIKE 165.
EGF-LIKE 166.
EGF-LIKE 167.
EGF-LIKE 168.
EGF-LIKE 169.
EGF-LIKE 170.
EGF-LIKE 171.
EGF-LIKE 172.
EGF-LIKE 173.
EGF-LIKE 174.
EGF-LIKE 175.
EGF-LIKE 176.
EGF-LIKE 177.
EGF-LIKE 178.
EGF-LIKE 179.
EGF-LIKE 180.
EGF-LIKE 181.
EGF-LIKE 182.
EGF-LIKE 183.
EGF-LIKE 184.
EGF-LIKE 185.
EGF-LIKE 186.
EGF-LIKE 187.
EGF-LIKE 188.
EGF-LIKE 189.
EGF-LIKE 190.
EGF-LIKE 191.
EGF-LIKE 192.
EGF-LIKE 193.
EGF-LIKE 194.
EGF-LIKE 195.
EGF-LIKE 196.
EGF-LIKE 197.
EGF-LIKE 198.
EGF-LIKE 199.
EGF-LIKE 200.
EGF-LIKE 201.
EGF-LIKE 202.
EGF-LIKE 203.
EGF-LIKE 204.
EGF-LIKE 205.
EGF-LIKE 206.
EGF-LIKE 207.
EGF-LIKE 208.
EGF-LIKE 209.
EGF-LIKE 210.
EGF-LIKE 211.
EGF-LIKE 212.
EGF-LIKE 213.
EGF-LIKE 214.
EGF-LIKE 215.
EGF-LIKE 216.
EGF-LIKE 217.
EGF-LIKE 218.
EGF-LIKE 219.
EGF-LIKE 220.
EGF-LIKE 221.
EGF-LIKE 222.
EGF-LIKE 223.
EGF-LIKE 224.
EGF-LIKE 225.
EGF-LIKE 226.
EGF-LIKE 227.
EGF-LIKE 228.
EGF-LIKE 229.
EGF-LIKE 230.
EGF-LIKE 231.
EGF-LIKE 232.
EGF-LIKE 233.
EGF-LIKE 234.
EGF-LIKE 235.
EGF-LIKE 236.
EGF-LIKE 237.
EGF-LIKE 238.
EGF-LIKE 239.
EGF-LIKE 240.
EGF-LIKE 241.
EGF-LIKE 242.
EGF-LIKE 243.
EGF-LIKE 244.
EGF-LIKE 245.
EGF-LIKE 246.
EGF-LIKE 247.
EGF-LIKE 248.
EGF-LIKE 249.
EGF-LIKE 250.
EGF-LIKE 251.
EGF-LIKE 252.
EGF-LIKE 253.
EGF-LIKE 254.
EGF-LIKE 255.
EGF-LIKE 256.
EGF-LIKE 257.
EGF-LIKE 258.
EGF-LIKE 259.
EGF-LIKE 260.
EGF-LIKE 261.
EGF-LIKE 262.
EGF-LIKE 263.
EGF-LIKE 264.
EGF-LIKE 265.
EGF-LIKE 266.
EGF-LIKE 267.
EGF-LIKE 268.
EGF-LIKE 269.
EGF-LIKE 270.
EGF-LIKE 271.
EGF-LIKE 272.
EGF-LIKE 273.
EGF-LIKE 274.
EGF-LIKE 275.
EGF-LIKE 276.
EGF-LIKE 277.
EGF-LIKE 278.
EGF-LIKE 279.
EGF-LIKE 280.
EGF-LIKE 281.
EGF-LIKE 282.
EGF-LIKE 283.
EGF-LIKE 284.
EGF-LIKE 285.
EGF-LIKE 286.
EGF-LIKE 287.
EGF-LIKE 288.
EGF-LIKE 289.
EGF-LIKE 290.
EGF-LIKE 291.
EGF-LIKE 292.
EGF-LIKE 293.
EGF-LIKE 294.
EGF-LIKE 295.
EGF-LIKE 296.
EGF-LIKE 297.
EGF-LIKE 298.
EGF-LIKE 299.
EGF-LIKE 300.
EGF-LIKE 301.
EGF-LIKE 302.
EGF-LIKE 303.
EGF-LIKE 304.
EGF-LIKE 305.
EGF-LIKE 306.
EGF-LIKE 307.
EGF-LIKE 308.
EGF-LIKE 309.
EGF-LIKE 310.
EGF-LIKE 311.
EGF-LIKE 312.
EGF-LIKE 313.
EGF-LIKE 314.
EGF-LIKE 315.
EGF-LIKE 316.
EGF-LIKE 317.
EGF-LIKE 318.
EGF-LIKE 319.
EGF-LIKE 320.
EGF-LIKE 321.
EGF-LIKE 322.
EGF-LIKE 323.
EGF-LIKE 324.
EGF-LIKE 325.
EGF-LIKE 326.
EGF-LIKE 327.
EGF-LIKE 328.
EGF-LIKE 329.
EGF-LIKE 330.
EGF-LIKE 331.
EGF-LIKE 332.
EGF-LIKE 333.
EGF-LIKE 334.
EGF-LIKE 335.
EGF-LIKE 336.
EGF-LIKE 337.
EGF-LIKE 338.
EGF-LIKE 339.
EGF-LIKE 340.
EGF-LIKE 341.
EGF-LIKE 342.
EGF-LIKE 343.
EGF-LIKE 344.
EGF-LIKE 345.
EGF-LIKE 346.
EGF-LIKE 347.
EGF-LIKE 348.
EGF-LIKE 349.
EGF-LIKE 350.
EGF-LIKE 351.
EGF-LIKE 352.
EGF-LIKE 353.
EGF-LIKE 354.
EGF-LIKE 355.
EGF-LIKE 356.
EGF-LIKE 357.
EGF-LIKE 358.
EGF-LIKE 359.
EGF-LIKE 360.
EGF-LIKE 361.
EGF-LIKE 362.
EGF-LIKE 363.
EGF-LIKE 364.
EGF-LIKE 365.
EGF-LIKE 366.
EGF-LIKE 367.
EGF-LIKE 368.
EGF-LIKE 369.
EGF-LIKE 370.
EGF-LIKE 371.
EGF-LIKE 372.
EGF-LIKE 373.
EGF-LIKE 374.
EGF-LIKE 375.
EGF-LIKE 376.
EGF-LIKE 377.
EGF-LIKE 378.
EGF-LIKE 379.
EGF-LIKE 380.
EGF-LIKE 381.
EGF-LIKE 382.
EGF-LIKE 383.
EGF-LIKE 384.
EGF-LIKE 385.
EGF-LIKE 386.
EGF-LIKE 387.
EGF-LIKE 388.
EGF-LIKE 389.
EGF-LIKE 390.
EGF-LIKE 391.
EGF-LIKE 392.
EGF-LIKE 393.
EGF-LIKE 394.
EGF-LIKE 395.
EGF-LIKE 396.
EGF-LIKE 397.
EGF-LIKE 398.
EGF-LIKE 399.
EGF-LIKE 400.
EGF-LIKE 401.
EGF-LIKE 402.
EGF-LIKE 403.
EGF-LIKE 404.
EGF-LIKE 405.
EGF-LIKE 406.
EGF-LIKE 407.
EGF-LIKE 408.
EGF-LIKE 409.
EGF-LIKE 410.
EGF-LIKE 411.
EGF-LIKE 412.
EGF-LIKE 413.
EGF-LIKE 414.
EGF-LIKE 415.
EGF-LIKE 416.
EGF-LIKE 417.
EGF-LIKE 418.
EGF-LIKE 419.
EGF-LIKE 420.
EGF-LIKE 421.
EGF-LIKE 422.
EGF-LIKE 423.
EGF-LIKE 424.
EGF-LIKE 425.
EGF-LIKE 426.
EGF-LIKE 427.
EGF-LIKE 428.
EGF-LIKE 429.
EGF-LIKE 430.
EGF-LIKE 431.
EGF-LIKE 432.
EGF-LIKE 433.
EGF-LIKE 434.
EGF-LIKE 435.
EGF-LIKE 436.
EGF-LIKE 437.
EGF-LIKE 438.
EGF-LIKE 439.
EGF-LIKE 440.
EGF-LIKE 441.
EGF-LIKE 442.
EGF-LIKE 443.
EGF-LIKE 444.
EGF-LIKE 445.
EGF-LIKE 446.
EGF-LIKE 447.
EGF-LIKE 448.
EGF-LIKE 449.
EGF-LIKE 450.
EGF-LIKE 451.
EGF-LIKE 452.
EGF-LIKE 453.
EGF-LIKE 454.
EGF-LIKE 455.
EGF-LIKE 456.
EGF-LIKE 457.
EGF-LIKE 458.
EGF-LIKE 459.
EGF-LIKE 460.
EGF-LIKE 461.
EGF-LIKE 462.
EGF-LIKE 463.
EGF-LIKE 464.
EGF-LIKE 465.
EGF-LIKE 466.
EGF-LIKE 467.
EGF-LIKE 468.
EGF-LIKE 469.
EGF-LIKE 470.
EGF-LIKE 471.
EGF-LIKE 472.
EGF-LIKE 473.
EGF-LIKE 474.
EGF-LIKE 475.
EGF-LIKE 476.
EGF-LIKE 477.
EGF-LIKE 478.
EGF-LIKE 479.
EGF-LIKE 480.
EGF-LIKE 481.
EGF-LIKE 482.
EGF-LIKE 483.
EGF-LIKE 484.
EGF-LIKE 485.
EGF-LIKE 486.
EGF-LIKE 487.
EGF-LIKE 488.
EGF-LIKE 489.
EGF-LIKE 490.
EGF-LIKE 491.
EGF-LIKE 492.
EGF-LIKE 493.
EGF-LIKE 494.
EGF-LIKE 495.
EGF-LIKE 496.
EGF-LIKE 497.
EGF-LIKE 498.
EGF-LIKE 499.
EGF-LIKE 500.
EGF-LIKE 501.
EGF-LIKE 502.
EGF-LIKE 503.
EGF-LIKE 504.
EGF-LIKE 505.
EGF-LIKE 506.
EGF-LIKE 507.
EGF-LIKE 508.
EGF-LIKE 509.
EGF-LIKE 510.
EGF-LIKE 511.
EGF-LIKE 512.
EGF-LIKE 513.
EGF-LIKE 514.
EGF-LIKE 515.
EGF-LIKE 516.
EGF-LIKE 517.
EGF-LIKE 518.
EGF-LIKE 519.
EGF-LIKE 520.
EGF-LIKE 521.
EGF-LIKE 522.
EGF-LIKE 523.
EGF-LIKE 524.
EGF-LIKE 525.
EGF-LIKE 526.
EGF-LIKE 527.
EGF-LIKE 528.
EGF-LIKE 529.
EGF-LIKE 530.
EGF-LIKE 531.
EGF-LIKE 532.
EGF-LIKE 533.
EGF-LIKE 534.
EGF-LIKE 535.
EGF-LIKE 536.
EGF-LIKE 537.
EGF-LIKE 538.
EGF-LIKE 539.
EGF-LIKE 540.
EGF-LIKE 541.
EGF-LIKE 542.
EGF-LIKE 543.
EGF-LIKE 544.
EGF-LIKE 545.
EGF-LIKE 546.
EGF-LIKE 547.
EGF-LIKE 548.
EGF-LIKE 549.
EGF-LIKE 550.
EGF-LIKE 551.
EGF-LIKE 552.
EGF-LIKE 553.
EGF-LIKE 554.
EGF-LIKE 555.
EGF-LIKE 556.
EGF-LIKE 557.
EGF-LIKE 558.
EGF-LIKE 559.
EGF-LIKE 560.
EGF-LIKE 561.
EGF-LIKE 562.
EGF-LIKE 563.
EGF-LIKE 564.
EGF-LIKE 565.
EGF-LIKE 566.
EGF-LIKE 567.
EGF-LIKE 568.
EGF-LIKE 569.
EGF-LIKE 570.
EGF-LIKE 571.
EGF-LIKE 572.
EGF-LIKE 573.
EGF-LIKE 574.
EGF-LIKE 575.
EGF-LIKE 576.
EGF-LIKE 577.
EGF-LIKE 578.
EGF-LIKE 579.
EGF-LIKE 580.
EGF-LIKE 581.
EGF-LIKE 582.
EGF-LIKE 583.
EGF-LIKE 584.
EGF-LIKE 585.
EGF-LIKE 586.
EGF-LIKE 587.
EGF-LIKE 588.
EGF-LIKE 589.
EGF-LIKE 590.
EGF-LIKE 591.
EGF-LIKE 592.
EGF-LIKE 593.
EGF-LIKE 594.
EGF-LIKE 595.
EGF-LIKE 596.
EGF-LIKE 597.
EGF-LIKE 598.
EGF-LIKE 599.
EGF-LIKE 600.
EGF-LIKE 601.
EGF-LIKE 602.
EGF-LIKE 603.
EGF-LIKE 604.
EGF-LIKE 605.
EGF-LIKE 606.
EGF-LIKE 607.
EGF-LIKE 608.
EGF-LIKE 609.
EGF-LIKE 610.
EGF-LIKE 611.
EGF-LIKE 612.
EGF-LIKE 613.
EGF-LIKE 614.
EGF-LIKE 615.
EGF-LIKE 616.
EGF-LIKE 617.
EGF-LIKE 618.
EGF-LIKE 619.
EGF-LIKE 620.
EGF-LIKE 621.
EGF-LIKE 622.
EGF-LIKE 623.
EGF-LIKE 624.
EGF-LIKE 625.
EGF-LIKE 626.
EGF-LIKE 627.
EGF-LIKE 628.
EGF-LIKE 629.
EGF-LIKE 630.
EGF-LIKE 631.
EGF-LIKE 632.
EGF-LIKE 633.
EGF-LIKE 634.
EGF-LIKE 635.
EGF-LIKE 636.
EGF-LIKE 637.
EGF-LIKE 638.
EGF-LIKE 639.
EGF-LIKE 640.
EGF-LIKE 641.
EGF-LIKE 642.
EGF-LIKE 643.
EGF-LIKE 644.
EGF-LIKE 645.
EGF-LIKE 646.
EGF-LIKE 647.
EGF-LIKE 648.
EGF-LIKE 649.
EGF-LIKE 650.
EGF-LIKE 651.
EGF-LIKE 652.
EGF-LIKE 653.
EGF-LIKE 654.
EGF-LIKE 655.
EGF-LIKE 656.
EGF-LIKE 657.
EGF-LIKE 658.
EGF-LIKE 659.
EGF-LIKE 660.
EGF-LIKE 661.
EGF-LIKE 662.
EGF-LIKE 663.
EGF-LIKE 664.
EGF-LIKE 665.
EGF-LIKE 666.
EGF-LIKE 667.
EGF-LIKE 668.
EGF-LIKE 669.
EGF-LIKE 670.
EGF-LIKE 671.
EGF-LIKE 672.
EGF-LIKE 673.
EGF-LIKE 674.
EGF-LIKE 675.
EGF-LIKE 676.
EGF-LIKE 677.
EGF-LIKE 678.
EGF-LIKE 679.
EGF-LIKE 680.
EGF-LIKE 681.
EGF-LIKE 682.
EGF-LIKE 683.
EGF-LIKE 684.
EGF-LIKE 685.
EGF-LIKE 686.
EGF-LIKE 687.
EGF-LIKE 688.
EGF-LIKE 689.
EGF-LIKE 690.
EGF-LIKE 691.
EGF-LIKE 692.
EGF-LIKE 693.
EGF-LIKE 694.
EGF-LIKE 695.
EGF-LIKE 696.
EGF-LIKE 697.
EGF-LIKE 698.
EGF-LIKE 699.
EGF-LIKE 700.
EGF-LIKE 701.
EGF-LIKE 702.
EGF-LIKE 703.
EGF-LIKE 704.
EGF-LIKE 705.
EGF-LIKE 706.
EGF-LIKE 707.
EGF-LIKE 708.
EGF-LIKE 709.
EGF-LIKE 710.
EGF-LIKE 711.
EGF-LIKE 712.
EGF-LIKE 713.
EGF-LIKE 714.
EGF-LIKE 715.
EGF-LIKE 716.
EGF-LIKE 717.
EGF-LIKE 718.
EGF-LIKE 719.
EGF-LIKE 720.
EGF-LIKE 721.
EGF-LIKE 722.
EGF-LIKE 723.
EGF-LIKE 724.
EGF-LIKE 725.
EGF-LIKE 726.
EGF-LIKE 727.
EGF-LIKE 728.
EGF-LIKE 729.
EGF-LIKE 730.
EGF-LIKE 731.
EGF-LIKE 732.
EGF-LIKE 733.
EGF-LIKE 734.
EGF-LIKE 735.
EGF-LIKE 736.
EGF-LIKE 737.
EGF-LIKE 738.
EGF-LIKE 739.
EGF-LIKE 740.
EGF-LIKE 741.
EGF-LIKE 742.
EGF-LIKE 743.
EGF-LIKE 744.
EGF-LIKE 745.
EGF-LIKE 746.
EGF-LIKE 747.
EGF-LIKE 748.
EGF-LIKE 749.
EGF-LIKE 750.
EGF-LIKE 751.
EGF-LIKE 752.
EGF-LIKE 753.
EGF-LIKE 754.
EGF-LIKE 755.
EGF-LIKE 756.
EGF-LIKE 757.
EGF-LIKE 758.
EGF-LIKE 759.
EGF-LIKE 760.
EGF-LIKE 761.
EGF-LIKE 762.
EGF-LIKE 763.
EGF-LIKE 764.
EGF-LIKE 765.
EGF-LIKE 766.
EGF-LIKE 767.
EGF-LIKE 768.
EGF-LIKE 769.
EGF-LIKE 770.
EGF-LIKE 771.
EGF-LIKE 772.
EGF-LIKE 773.
EGF-LIKE 774.
EGF-LIKE 775.
EGF-LIKE 776.
EGF-LIKE 777.
EGF-LIKE 778.
EGF-LIKE 779.
EGF-LIKE 780.
EGF-LIKE 781.
EGF-LIKE 782.
EGF-LIKE 783.
EGF-LIKE 784.
EGF-LIKE 785.
EGF-LIKE 786.
EGF-LIKE 787.
EGF-LIKE 788.
EGF-LIKE 789.
EGF-LIKE 790.
EGF-LIKE 791.
EGF-LIKE 792.
EGF-LIKE 793.
EGF-LIKE 794.
EGF-LIKE 795.
EGF-LIKE 796.
EGF-LIKE 797.
EGF-LIKE 798.
EGF-LIKE 799.
EGF-LIKE 800.
EGF-LIKE 801.
EGF-LIKE 802.
EGF-LIKE 803.
EGF-LIKE 804.
EGF-LIKE 805.
EGF-LIKE 806.
EGF-LIKE 807.
EGF-LIKE 808.
EGF-LIKE 809.
EGF-LIKE 810.
EGF-LIKE 811.
EGF-LIKE 812.
EGF-LIKE 813.
EGF-LIKE 814.
EGF-LIKE 815.
EGF-LIKE 816.
EGF-LIKE 817.
EGF-LIKE 818.
EGF-LIKE 819.
EGF-LIKE 820.
EGF-LIKE 821.
EGF-LIKE 822.
EGF-LIKE 823.
EGF-LIKE 824.
EGF-LIKE 825.
EGF-LIKE 826.
EGF-LIKE 827.
EGF-LIKE 828.
EGF-LIKE 829.
EGF-LIKE 830.
EGF-LIKE 831.
EGF-LIKE 832.
EGF-LIKE 833.
EGF-LIKE 834.
EGF-LIKE 835.
EGF-LIKE 836.
EGF-LIKE 837.
EGF-LIKE 838.
EGF-LIKE 839.
EGF-LIKE 840.
EGF-LIKE 841.
EGF-LIKE 842.
EGF-LIKE 843.
EGF-LIKE 844.
EGF-LIKE 845.
EGF-LIKE 846.
EGF-LIKE 847.
EGF-LIKE 848.
EGF-LIKE 849.
EGF-LIKE 850.
EGF-LIKE 851.
EGF-LIKE 852.
EGF-LIKE 853.
EGF-LIKE 854.
EGF-LIKE 855.
EGF-LIKE 856.
EGF-LIKE 857.
EGF-LIKE 858.
EGF-LIKE 859.
EGF-LIKE 860.
EGF-LIKE 861.
EGF-LIKE 862.
EGF-LIKE 863.
EGF-LIKE 864.
EGF-LIKE 865.
EGF-LIKE 866.
EGF-LIKE 867.
EGF-LIKE 868.
EGF-LIKE 869.
EGF-LIKE 870.
EGF-LIKE 871.
EGF-LIKE 872.
EGF-LIKE 873.
EGF-LIKE 874.
EGF-LIKE 875.
EGF-LIKE 876.
EGF-LIKE 877.
EGF-LIKE 878.
EGF-LIKE 879.
EGF-LIKE 880.
EGF-LIKE 881.
EGF-LIKE 882.
EGF-LIKE 883.
EGF-LIKE 884.
EGF-LIKE 885.
EGF-LIKE 886.
EGF-LIKE 887.
EGF-LIKE 888.
EGF-LIKE 889.
EGF-LIKE 890.
EGF-LIKE 891.
EGF-LIKE 892.
EGF-LIKE 893.
EGF-LIKE 894.
EGF-LIKE 895.
EGF-LIKE 896.
EGF-LIKE 897.
EGF-LIKE 898.
EGF-LIKE 899.
EGF-LIKE 900.
EGF-LIKE 901.
EGF-LIKE 902.
EGF-LIKE 903.
EGF-LIKE 904.
EGF-LIKE 905.
EGF-LIKE 906.
EGF-LIKE 907.
EGF-LIKE 908.
EGF-LIKE 909.
EGF-LIKE 910.
EGF-LIKE 911.
EGF-LIKE 912.
EGF-LIKE 913.
EGF-LIKE 914.
EGF-LIKE 915.
EGF-LIKE 916.
EGF-LIKE 917.
EGF-LIKE 918.
EGF-LIKE 919.
EGF-LIKE 920.
EGF-LIKE 921.
EGF-LIKE 922.
EGF-LIKE 923.
EGF-LIKE 924.
EGF-LIKE 925.
EGF-LIKE 926.
EGF-LIKE 927.
EGF-LIKE 928.
EGF-LIKE 929.
EGF-LIKE 930.
EGF-LIKE 931.
EGF-LIKE 932.
EGF-LIKE 933.
EGF-LIKE 934.
EGF-LIKE 935.
EGF-LIKE 936.
EGF-LIKE 937.
EGF-LIKE 938.
EGF-LIKE 939.
EGF-LIKE 940.
EGF-LIKE 941.
EGF-LIKE 942.
EGF-LIKE 943.
EGF-LIKE 944.
EGF-LIKE 945.
EGF-LIKE 946.
EGF-LIKE 947.
EGF-LIKE 948.
EGF-LIKE 949.
EGF-LIKE 950.
EGF-LIKE 951.
EGF-LIKE 952.
EGF-LIKE 953.
EGF-LIKE 954.
EGF-LIKE 955.
EGF-LIKE 956.
EGF-LIKE 957.
EGF-LIKE 958.
EGF-LIKE 959.
EGF-LIKE 960.
EGF-LIKE 961.
EGF-LIKE 962.
EGF-LIKE 963.
EGF-LIKE 964.
EGF-LIKE 965.
EGF-LIKE 966.
EGF-LIKE 967.
EGF-LIKE 968.
EGF-LIKE 969.
EGF-LIKE 970.
EGF-LIKE 971.
EGF-LIKE 972.
EGF-LIKE 973.
EGF
```

FT	DISULFID	111	127	BY SIMILARITY
FT	DISULFID	129	138	BY SIMILARITY
FT	DISULFID	144	155	BY SIMILARITY
FT	DISULFID	149	164	BY SIMILARITY
FT	DISULFID	166	175	BY SIMILARITY
FT	DISULFID	182	195	BY SIMILARITY
FT	DISULFID	189	204	BY SIMILARITY
FT	DISULFID	206	215	BY SIMILARITY
FT	DISULFID	222	233	BY SIMILARITY
FT	DISULFID	227	243	BY SIMILARITY
FT	DISULFID	245	254	BY SIMILARITY
FT	DISULFID	261	272	BY SIMILARITY
FT	DISULFID	266	281	BY SIMILARITY
FT	DISULFID	283	292	BY SIMILARITY
FT	DISULFID	299	312	BY SIMILARITY
FT	DISULFID	306	321	BY SIMILARITY
FT	DISULFID	323	332	BY SIMILARITY
FT	DISULFID	339	350	BY SIMILARITY
FT	DISULFID	344	359	BY SIMILARITY
FT	DISULFID	361	370	BY SIMILARITY
FT	DISULFID	376	387	BY SIMILARITY
FT	DISULFID	381	398	BY SIMILARITY
FT	DISULFID	400	409	BY SIMILARITY
FT	DISULFID	416	429	BY SIMILARITY
FT	DISULFID	423	438	BY SIMILARITY
FT	DISULFID	440	449	BY SIMILARITY
FT	DISULFID	456	467	BY SIMILARITY
FT	DISULFID	461	476	BY SIMILARITY
FT	DISULFID	478	487	BY SIMILARITY
FT	DISULFID	494	505	BY SIMILARITY
FT	DISULFID	499	514	BY SIMILARITY
FT	DISULFID	516	525	BY SIMILARITY
FT	DISULFID	532	543	BY SIMILARITY
FT	DISULFID	537	552	BY SIMILARITY
FT	DISULFID	554	563	BY SIMILARITY
FT	DISULFID	570	580	BY SIMILARITY
FT	DISULFID	575	589	BY SIMILARITY
FT	DISULFID	591	600	BY SIMILARITY
FT	DISULFID	607	618	BY SIMILARITY
FT	DISULFID	612	627	BY SIMILARITY
FT	DISULFID	629	638	BY SIMILARITY
FT	DISULFID	645	655	BY SIMILARITY
FT	DISULFID	650	664	BY SIMILARITY
FT	DISULFID	666	675	BY SIMILARITY
FT	DISULFID	682	693	BY SIMILARITY
FT	DISULFID	687	702	BY SIMILARITY
FT	DISULFID	704	713	BY SIMILARITY
FT	DISULFID	720	730	BY SIMILARITY
FT	DISULFID	725	739	BY SIMILARITY
FT	DISULFID	741	750	BY SIMILARITY
FT	DISULFID	757	768	BY SIMILARITY
FT	DISULFID	762	777	BY SIMILARITY
FT	DISULFID	779	788	BY SIMILARITY
FT	DISULFID	795	806	BY SIMILARITY
FT	DISULFID	800	815	BY SIMILARITY
FT	DISULFID	817	826	BY SIMILARITY
FT	DISULFID	833	844	BY SIMILARITY
FT	DISULFID	838	855	BY SIMILARITY
FT	DISULFID	857	866	BY SIMILARITY
FT	DISULFID	873	884	BY SIMILARITY
FT	DISULFID	895	904	BY SIMILARITY
FT	DISULFID	911	922	BY SIMILARITY
FT	DISULFID	916	931	BY SIMILARITY
FT	DISULFID	933	942	BY SIMILARITY
FT	DISULFID	987	998	BY SIMILARITY
FT	DISULFID	992	1007	BY SIMILARITY
FT	DISULFID	1009	1018	BY SIMILARITY
FT	DISULFID	1025	1036	BY SIMILARITY
FT	DISULFID	1030	1045	BY SIMILARITY
FT	DISULFID	1047	1056	BY SIMILARITY
FT	DISULFID	1063	1074	BY SIMILARITY

[illegible]

RESULT	24		
PRTS_RABIT			
ID	PRTS_RABIT	STANDARD:	PRT: 646 AA.
AC	P96118.		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Vitamin K-dependent protein S precursor (Fragment).		
GN	PROS1 OR PROS.		
OS	Oryctolagus cuniculus (Rabbit).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
OX	NCBI_TaxID=9986;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=94039141; PubMed=8223642;		
RA	He X., Dahlbaeck B.;		
RT	"Molecular cloning, expression and functional characterization of		
RT	rabbit anticoagulant vitamin K-dependent protein S.";		
RL	Eur. J. Biochem. 217:857-865(1993).		
CC	- FUNCTION: ANTI COAGULANT PLASMA PROTEIN; IT IS A COFACTOR TO		
CC	ACTIVATED PROTEIN C IN THE DEGRADATION OF COAGULATION FACTORS VA		
CC	AND VIIIA. IT HELP TO PREVENT COAGULATION AND STIMULATING		
CC	FIBRINOLYSIS.		
CC	- FUNCTION: PROTEIN S INTERACTS WITH C4B-BINDING PROTEIN, A		
CC	REGULATOR OF THE COMPLEX SYSTEM. IN RABBIT PLASMA HOWEVER, PROTEIN		
CC	S APPEARS TO BE PRESENT ONLY IN FREE FORM.		
CC	- SUBCELLULAR LOCATION: Extracellular.		
CC	- TISSUE SPECIFICITY: PLASMA.		
CC	- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.		
CC	- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation on		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).		
CC	-----		
DR	EMBL; Z26485; CAAB1259.1; -		
DR	HSSP; P00740; ICFH.		
DR	InterPro; IPR000152; Asx_hydroxyl.		
DR	InterPro; IPR000561; EGF-like.		

FT DOMAIN 489 525 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 527 563 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 565 600 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 602 638 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 640 675 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 677 713 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 715 750 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 752 788 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 790 826 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 828 866 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 868 904 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 906 942 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 944 980 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 982 1018 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1020 1056 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1058 1094 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1096 1142 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1144 1180 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1182 1218 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1220 1264 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1266 1304 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1306 1346 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1347 1383 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1386 1424 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 1441 1478 LIN/NOTCH 1.
FT REPEAT 1479 1520 LIN/NOTCH 2.
FT REPEAT 1521 1560 LIN/NOTCH 3.
FT REPEAT 1562 1599 ANK 1.
FT REPEAT 1599 1636 ANK 2.
FT REPEAT 1636 1673 ANK 3.
FT REPEAT 1673 1710 ANK 4.
FT REPEAT 1710 1747 ANK 5.
FT REPEAT 1747 1784 ANK 6.
FT DISULFID 22 35 BY SIMILARITY.
FT DISULFID 35 48 BY SIMILARITY.
FT DISULFID 48 61 BY SIMILARITY.
FT DISULFID 61 74 BY SIMILARITY.
FT DISULFID 74 87 BY SIMILARITY.
FT DISULFID 87 98 BY SIMILARITY.
FT DISULFID 98 111 BY SIMILARITY.
FT DISULFID 111 128 BY SIMILARITY.
FT DISULFID 128 139 BY SIMILARITY.
FT DISULFID 139 156 BY SIMILARITY.
FT DISULFID 156 165 BY SIMILARITY.
FT DISULFID 165 176 BY SIMILARITY.
FT DISULFID 176 194 BY SIMILARITY.
FT DISULFID 194 203 BY SIMILARITY.
FT DISULFID 203 214 BY SIMILARITY.
FT DISULFID 214 232 BY SIMILARITY.
FT DISULFID 232 242 BY SIMILARITY.
FT DISULFID 242 253 BY SIMILARITY.
FT DISULFID 253 271 BY SIMILARITY.
FT DISULFID 271 280 BY SIMILARITY.
FT DISULFID 280 291 BY SIMILARITY.
FT DISULFID 291 311 BY SIMILARITY.
FT DISULFID 311 320 BY SIMILARITY.
FT DISULFID 320 331 BY SIMILARITY.
FT DISULFID 331 349 BY SIMILARITY.
FT DISULFID 349 358 BY SIMILARITY.
FT DISULFID 358 369 BY SIMILARITY.
FT DISULFID 369 386 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 397 408 BY SIMILARITY.
FT DISULFID 408 428 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 437 448 BY SIMILARITY.
FT DISULFID 448 466 BY SIMILARITY.
FT DISULFID 466 475 BY SIMILARITY.
FT DISULFID 475 486 BY SIMILARITY.
FT DISULFID 486 504 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 513 524 BY SIMILARITY.
FT DISULFID 524 542 BY SIMILARITY.

FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 553 562 BY SIMILARITY.
FT DISULFID 569 579 BY SIMILARITY.
FT DISULFID 574 588 BY SIMILARITY.
FT DISULFID 590 599 BY SIMILARITY.
FT DISULFID 606 617 BY SIMILARITY.
FT DISULFID 621 626 BY SIMILARITY.
FT DISULFID 628 637 BY SIMILARITY.
FT DISULFID 644 654 BY SIMILARITY.
FT DISULFID 649 663 BY SIMILARITY.
FT DISULFID 665 674 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.
FT DISULFID 686 701 BY SIMILARITY.
FT DISULFID 703 712 BY SIMILARITY.
FT DISULFID 719 729 BY SIMILARITY.
FT DISULFID 724 738 BY SIMILARITY.
FT DISULFID 740 749 BY SIMILARITY.
FT DISULFID 756 767 BY SIMILARITY.
FT DISULFID 761 776 BY SIMILARITY.
FT DISULFID 778 787 BY SIMILARITY.
FT DISULFID 794 805 BY SIMILARITY.
FT DISULFID 799 814 BY SIMILARITY.
FT DISULFID 816 825 BY SIMILARITY.
FT DISULFID 832 843 BY SIMILARITY.
FT DISULFID 837 854 BY SIMILARITY.
FT DISULFID 856 865 BY SIMILARITY.
FT DISULFID 872 883 BY SIMILARITY.
FT DISULFID 877 892 BY SIMILARITY.
FT DISULFID 894 903 BY SIMILARITY.
FT DISULFID 910 921 BY SIMILARITY.
FT DISULFID 915 930 BY SIMILARITY.
FT DISULFID 932 941 BY SIMILARITY.
FT DISULFID 946 955 BY SIMILARITY.
FT DISULFID 966 977 BY SIMILARITY.
FT DISULFID 981 1006 BY SIMILARITY.
FT DISULFID 1008 1017 BY SIMILARITY.
FT DISULFID 1024 1035 BY SIMILARITY.
FT DISULFID 1029 1044 BY SIMILARITY.
FT DISULFID 1046 1055 BY SIMILARITY.
FT DISULFID 1062 1073 BY SIMILARITY.
FT DISULFID 1067 1082 BY SIMILARITY.
FT DISULFID 1084 1093 BY SIMILARITY.
FT DISULFID 1100 1121 BY SIMILARITY.
FT DISULFID 1115 1130 BY SIMILARITY.
FT DISULFID 1132 1141 BY SIMILARITY.
FT DISULFID 1148 1159 BY SIMILARITY.
FT DISULFID 1153 1168 BY SIMILARITY.

Query Match 11.1%; Score 223; DB 1; Length 2524;
Best Local Similarity 23.5%; Pred. No. 5.1e-09;
Matches 68; Conservative 20; Mismatches 101; Indels 100; Gaps 14;

Qy 126 VC-CSPGYTPDCL---ACGGGSGRPCSGNGHSGGDSRGDGSRCRHMGYGFPLCTD- 179
Db 164 ICKCPFGHGATCKODINEC---SQNPCKNGQCINE--FGSYRCTCQNRFTGHNCDP 217

Qy 180 -----CMGDFSSLANETHSTCTA-----CDSCSKRC-----SGTJN 211
Db 218 YVPCNPSPCLNNGTCRQTDDBTSYDCTCLPFGSGQNCSEENIDCPNNCRNGSTCYDGVNT 277

Qy 212 RDCGCEVGVWYLDGAGCVNVDCAEPPCSAOPCKNANGSYTC-----EBCDSS- 262
Db 278 YNC-QCPDPWT-GQYCTEDVDGECQLMPPNACQNGCTHMTYGGYNCVCVNGWTGECSENI 335

Qy 263 --CVGCTGEGPNCKE-----CISGVAREHGQC----- 288
Db 336 DDCANAAHSGATCHDRVASFYCECPHGRGTGLCHLDNACISNPNESNCDTPNPNVNGKA 395

Qy 289 -----ADVDECSIAEKTGVKKNKNCVNTFGSYCYVCPDGF 323
Db 396 ICTCPPGYTGPAACNNDVDECSLGANPC--EHGRCNTNLTGSRQNCNPDGF 443

RESULT 27

ID	EMBL	HUMAN	STANDARD:	PRO:	886 AA.
AC	014246:				
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	01-MAR-2002	(Rel. 41, Last annotation update)			
DE	Cell surface glycoprotein EMR1 precursor (EMR1 hormone receptor).				
GN	EMR1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Euteeria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=95324926; PubMed=7601460;				
RA	Baud V., Chissoc S.L., Viegas-Pequignot E., Dilion S., N'Guyen V.C.,				
RA	Roe B.A., Liptinski M.,				
RT	"EMR1, an unusual member in the family of hormone receptors with				
RL	seven transmembrane segments."				
CC	Genomics 26:334-344(1995).				
CC	-1- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR				
CC	AN INTERACTION WITH A PROTEIN LIGAND.				
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.				
CC	-1- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN				
CC	PERIPHERAL BLOOD MONONUCLEAR CELLS.				
CC	-1- PTM: N- AND O-GLYCOSYLATED; (POTENTIAL).				
CC	-1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.				
CC	-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.				
CC	-1- SIMILARITY: CONTAINS 1 GPS DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X81479; CA457232.1; -.				
DR	HSSP; P35555; IEMN.				
DR	MIM; 600493;				
DR	InterPro: IPR000152; Asx_hydroxyl.				
DR	InterPro: IPR000561; EGF-like.				
DR	InterPro: IPR001881; EGF Ca.				
DR	InterPro: IPR008832; GPCR_secretin.				
DR	InterPro: IPR000203; PKD_cys_tich.				
DR	Pfam; PF00002; 7tm_2; 1.				
DR	Pfam; PF00008; EGF; 5.				
DR	Pfam; PF01825; GPS; 1.				
DR	PRINTS; PR01128; EMR1HORMONER.				
DR	SMART; SM00179; EGF_CA; 5.				
DR	SMART; SM00001; EGF-like; 1.				
DR	SMART; SM00303; GPS; 1.				
DR	PROSITE; PS00010; ASX_HYDROXYL; 6.				
DR	PROSITE; PS01186; EGF_2; 2.				
DR	PROSITE; PS01187; EGF_CA; 5.				
DR	PROSITE; PS02221; GPS; 1.				
DR	PROSITE; PS00650; G-PROTEIN_RECIP_F2_2; 1.				
DR	PROSITE; PS02621; G-PROTEIN_RECIP_F2_4; 1.				
KW	G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;				
KW	EGF-like domain; Repeat; Signal.				
FT	SIGNAL	1	17	POTENTIAL.	
FT	CHAIN	18	886	CELL SURFACE GLYCOPROTEIN EMR1.	
FT	DOMAIN	18	599	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM	600	627	POTENTIAL.	
FT	DOMAIN	628	634	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM	635	656	POTENTIAL.	
FT	DOMAIN	657	666	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM	667	690	POTENTIAL.	
FT	DOMAIN	691	709	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM	710	731	POTENTIAL.	
FT	DOMAIN	732	747	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM	748	776	POTENTIAL.	

Query Match	Best Local Similarity	Score	DB 1	Length	886:
Matches	79:	Conservative	20:	Mismatches	95:
				Indels	89:
					Gaps
130	PETGYPRDCLACGGGSGQRPCSGNGHCSGGDSROGGDSRCRMGYQGPLCTDCM---	DGYRS	186		
Db	74	PCVPRCKDIDEC--SQSPROPQGNSSCK--NLSGRKYSCLDGFSSPTGNDWVPGKGNFS	129		
187	SLRNETHSICTCADSCSKCKSGLTNRDGE-----	CEVGVNLDGACVVD	232		
Db	130	-----CTDINE---C--LTISVCPESHSDCVNMSGSYSCQGVFISRNSTCEDVN	174		
233	ECAAEPSPCSAOPCFKNANGSYTC-----	EECDSSC--V	264		
Db	175	EC-ADPRACPEHATCNNTNMGVNSYSCPNCFESSSGHLSCGGLKASCEIDDECTEMCPINS	233		
265	GCTGGSPGN--CKEELSGYAREHQ-----	CADVDECSLAETCYRKNECNYTFG	313		
Db	234	TCTNT-PSISYF-TCHPGAPSSQINFITDQGEVCRDIDECQDPSYC-GPNSICTNALG	290		
314	SVYCVCPDGFETEDACVPAEAFEGESPRQ---	LPSEBPL	353		
Db	291	SYSCGCIYGFH-----PNEPESQKDGNFSCQRIWLFCKREY	326		

AC	000918;1993 (Rel. 26, Created)
DT	01-JUL-1993 (Rel. 26, Last sequence update)
DT	01-JUL-1993 (Rel. 26, Last annotation update)
DE	Latent transforming growth factor beta binding protein 1 precursor
DE	(Transforming growth factor beta-1 binding protein 1) (TGF-beta1-BP-
GN	LTPB1).
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Muriinae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX	MEDLINE=91062373; PubMed=2247454;
RA	Tsuji T., Okada F., Yamaguchi K., Nakamura T.;
RT	"Molecular cloning of the large subunit of transforming growth factor
RT	type beta masking protein and expression of the mRNA in various rat
RL	tissues".
RC	Proc. Natl. Acad. Sci. U.S.A. 87:8835-8839(1990).
CC	-I- SUBUNIT: THE LARGE LATENT COMPLEX OF TGF-BETA1 FROM PLATELETS IS
CC	COMPOSED OF THE TGF-BETA1 MOLECULE NONCOVALENTLY ASSOCIATED WITH
CC	A MASKING PROTEIN CONSISTING OF A DISULFIDE-BONDED COMPLEX OF A
CC	DIMER OF THE N-TERMINAL PROPEPTIDE OF THE TGF-BETA1 PRECURSOR AND
CC	A THIRD COMPONENT DENOTED TGF-BETA1-BP (OR MP LARGE SUBUNIT).
CC	TGF-BETA1-BP DOES NOT BIND DIRECTLY TO ACTIVE TGF-BETA1.
CC	-I- SIMILARITY: CONTAINS 18 EGF-LIKE DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
DR	EMBL; M55431; AAAA4235.1; -.
DR	PIR; A38261; A38261.
DR	HSSP; P35555; IEKN.
DR	InterPro; IPR0000152; Asx_hydroxy1.
DR	InterPro; IPR0000561; EGF_1like.
DR	InterPro; IPR001881; EGF_CA.
DR	InterPro; IPR002212; TB.
DR	Pfam; PF00008; EGF; 16.
DR	Pfam; PF00683; TB; 4.
DR	SMART; SMO0179; EGF_CA; 13.
DR	SMART; SMO0001; EGF_1like; 5.
DR	PROSITE; PS00010; ASX_HYDROXYL; 13.
DR	PROSITE; PS00022; EGF_1; 2.
DR	PROSITE; PS01186; EGF_2; 10.
DR	PROSITE; PS01187; EGF_CA; 15.
KW	Growth factor binding; Repeat; EGF-like domain; Hydroxylation; Signal.
FT	SIGNAL
FT	PROPEP
FT	CHAIN
FT	1 21 736
FT	737 1577
FT	SITE 734 736
FT	SITE 1575 1577
FT	PROPEP 1578 1712
FT	DOMAIN 181 213
FT	DOMAIN 391 423
FT	REPEAT 551 604
FT	DOMAIN 618 658
FT	REPEAT 671 721
FT	DOMAIN 865 906
FT	DOMAIN 907 948
FT	DOMAIN 949 989
FT	DOMAIN 990 1029
FT	DOMAIN 1030 1070
FT	DOMAIN 1071 1111
FT	DOMAIN 1112 1152
FT	DOMAIN 1153 1193
FT	DOMAIN 1194 1235

FT	DOMAIN	1236	1277	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1278	1320	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1340	1392	INTERNAL REPEAT 3
FT	DOMAIN	1415	1457	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1458	1498	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL)
FT	REPEAT	1517	1568	INTERNAL REPEAT 4
FT	DOMAIN	1612	1652	EGF-LIKE 17
FT	DOMAIN	1653	1697	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL)
FT	DISULFID	185	195	BY SIMILARITY.
FT	DISULFID	189	201	BY SIMILARITY.
FT	DISULFID	203	212	BY SIMILARITY.
FT	DISULFID	395	405	BY SIMILARITY.
FT	DISULFID	399	411	BY SIMILARITY.
FT	DISULFID	413	422	BY SIMILARITY.
FT	DISULFID	622	633	BY SIMILARITY.
FT	DISULFID	628	642	BY SIMILARITY.
FT	DISULFID	644	657	BY SIMILARITY.
FT	DISULFID	869	881	BY SIMILARITY.
FT	DISULFID	876	890	BY SIMILARITY.
FT	DISULFID	892	905	BY SIMILARITY.
FT	DISULFID	911	923	BY SIMILARITY.
FT	DISULFID	918	932	BY SIMILARITY.
FT	DISULFID	934	947	BY SIMILARITY.
FT	DISULFID	953	964	BY SIMILARITY.
FT	DISULFID	959	973	BY SIMILARITY.
FT	DISULFID	976	988	BY SIMILARITY.
FT	DISULFID	994	1005	BY SIMILARITY.
FT	DISULFID	1000	1014	BY SIMILARITY.
FT	DISULFID	1017	1028	BY SIMILARITY.
FT	DISULFID	1034	1045	BY SIMILARITY.
FT	DISULFID	1040	1054	BY SIMILARITY.
FT	DISULFID	1056	1069	BY SIMILARITY.
FT	DISULFID	1075	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1110	BY SIMILARITY.
FT	DISULFID	1116	1127	BY SIMILARITY.
FT	DISULFID	1122	1136	BY SIMILARITY.
FT	DISULFID	1138	1151	BY SIMILARITY.
FT	DISULFID	1157	1169	BY SIMILARITY.
FT	DISULFID	1164	1178	BY SIMILARITY.
FT	DISULFID	1180	1192	BY SIMILARITY.
FT	DISULFID	1198	1210	BY SIMILARITY.
FT	DISULFID	1204	1219	BY SIMILARITY.
FT	DISULFID	1221	1234	BY SIMILARITY.
FT	DISULFID	1240	1252	BY SIMILARITY.
FT	DISULFID	1246	1261	BY SIMILARITY.
FT	DISULFID	1263	1276	BY SIMILARITY.
FT	DISULFID	1282	1294	BY SIMILARITY.
FT	DISULFID	1289	1303	BY SIMILARITY.
FT	DISULFID	1305	1319	BY SIMILARITY.
FT	DISULFID	1419	1432	BY SIMILARITY.
FT	DISULFID	1427	1441	BY SIMILARITY.
FT	DISULFID	1443	1456	BY SIMILARITY.
FT	DISULFID	1462	1473	BY SIMILARITY.
FT	DISULFID	1468	1482	BY SIMILARITY.
FT	DISULFID	1484	1497	BY SIMILARITY.
FT	DISULFID	1616	1627	BY SIMILARITY.
FT	DISULFID	1622	1636	BY SIMILARITY.
FT	DISULFID	1638	1651	BY SIMILARITY.
FT	DISULFID	1657	1672	BY SIMILARITY.
FT	DISULFID	1687	1681	BY SIMILARITY.
FT	DISULFID	1683	1696	BY SIMILARITY.
FT	CARBOHYD	339	339	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	370	370	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	416	416	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	612	612	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1042	1042	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1242	1242	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1357	1357	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQ	SEQUENCE	1712	186598	MM; 16550BCEAA691FDJ34 CRC64;

Query Match

11.187

Score 222; DB 1; Length 1712;


```

FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 660 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 867 BY SIMILARITY.
FT DISULFID 874 885 BY SIMILARITY.
FT DISULFID 879 894 BY SIMILARITY.
FT DISULFID 896 905 BY SIMILARITY.
FT DISULFID 912 923 BY SIMILARITY.
FT DISULFID 917 932 BY SIMILARITY.
FT DISULFID 934 943 BY SIMILARITY.
FT DISULFID 988 999 BY SIMILARITY.
FT DISULFID 993 1008 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.
FT DISULFID 1026 1037 BY SIMILARITY.

```

```

FT DISULFID 1031 1046 BY SIMILARITY.
FT DISULFID 1048 1057 BY SIMILARITY.
FT DISULFID 1064 1075 BY SIMILARITY.

Query Match 11.1%; Score 222; DB 1; Length 2444;
Best Local Similarity 24.6%; Pred. No. 5.9e-09;
Matches 75; Conservative 26; Mismatches 104; Indels 100; Gaps 15;

QY 125 KVCSPGTY-GPDLACOGSGSPGSGNHCS-GDSSRGDGSRCRMGYOGLCTDCMD 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 EACVCGGATVGPQCDPNPCLSPCKMAGTCHVDKRGVADYACSCALFSGPLCLTPD 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 GYFSSLRNE-----THSICTADESCKTSGLTRNDGCE---CEWGVLDGACV----- 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 N-ACLTNCRNGTODLLTFEYKCRCPGWSGKSCQOADPCASPNANGGCLPFES 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 -----DVECAAEPPPCSAOFCRKNANSYTC----- 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 YICHPSPFHGPTCRQDVNDCGKPRLCRRHGGTCHNEVGSYRCVCRATHTGPCRPPYV 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 -----EECDSSCV-GCTGEG-----PG-NCKE---CISGYANE 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 CSPSPQNGTCKPTGVDYHEC--ACLPFTGQNCENENIDDCRGNNCKNGACVDSVNTY 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 HGQC-----ADVDECSIAEKTQVRKNENCVTPGSIYVCVCPDGF-----EETEDA 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 NCPCPPEMTGYCTEDVDQCQMPNAC--ONGGCHHTHGVCVNGWTGDECSNID 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 330 CVPPA 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 CASAA 343

RESULT 31
NIDO_HUMAN
ID NIDO_HUMAN STANDARD; PRT; 1247 AA.
AC P14543; Q14942;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nidogen precursor (Entactin).
GN NID.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90091745; PubMed=2574658;
RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,
  Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;
RT "Human nidogen: complete amino acid sequence and structural domains
  deduced from cDNAs, and evidence for polymorphism of the gene.";
RL DNA 8:581-594(1989).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=96044428; PubMed=7557988;
RX Zimmermann K., Holschen S., Hafner M., Nischt R.;
RT "Genomic sequences and structural organization of the human nidogen
  gene (NID).";
RL Genomics 27:245-250(1995).
[3]
RN SEQUENCE OF 667-1247 FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=89270475; PubMed=2471408;
RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Passage E., Weil D.,
  Timpl R., Chu M.L., Uitto J.;
RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the
  gene to chromosome 1q43.";
RL Am. J. Hum. Genet. 44:876-885(1989).
CC -!- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN
  BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-

```

```

CC EXTRACELLULAR MATRIX INTERACTIONS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
CC -1- PTM: N- AND O-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR WYTD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC used by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M30269; AAA59932.1; -.
DR EMBL, X82245; CAAS57709.1; JOINED.
DR EMBL, X84819; CAAS57709.1; JOINED.
DR EMBL, X84820; CAAS57709.1; JOINED.
DR EMBL, X84821; CAAS57709.1; JOINED.
DR EMBL, X84822; CAAS57709.1; JOINED.
DR EMBL, X84823; CAAS57709.1; JOINED.
DR EMBL, X84824; CAAS57709.1; JOINED.
DR EMBL, X84825; CAAS57709.1; JOINED.
DR EMBL, X84826; CAAS57709.1; JOINED.
DR EMBL, X84827; CAAS57709.1; JOINED.
DR EMBL, X84828; CAAS57709.1; JOINED.
DR EMBL, X84829; CAAS57709.1; JOINED.
DR EMBL, X84830; CAAS57709.1; JOINED.
DR EMBL, X84831; CAAS57709.1; JOINED.
DR EMBL, X84832; CAAS57709.1; JOINED.
DR EMBL, X84833; CAAS57709.1; JOINED.
DR EMBL, X84834; CAAS57709.1; JOINED.
DR EMBL, X84835; CAAS57709.1; JOINED.
DR EMBL, X84836; CAAS57709.1; JOINED.
DR EMBL, X84837; CAAS57709.1; JOINED.
DR EMBL, M27445; AAA57261.1; -.
DR PIR, A33322; MIMHUND.
DR HSSP, P07204; IADX.
DR MIM, J131390; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR000033; Ldl_receptor_rep.
DR InterPro: IPR003866; Nidogen_ext.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam, PF00008; EGF_6.
DR Pfam, PF00058; ldl_recept_b_3.
DR Pfam, PF00086; thyroglobulin_1; 1.
DR SMART, SM00179; EGF_CA_2.
DR SMART, SM00001; EGF_Like; 4.
DR SMART, SM00135; LY; 5.
DR SMART, SM00539; NIDO_1.
DR SMART, SM00211; TY; 1.
DR PROSITE, PS00010; ASX_HYDROXYL; 3.
DR PROSITE, PS00484; THYROGLOBULIN_1; 1.
DR PROSITE, PS00022; EGF_1; 1.
DR PROSITE, PS01186; EGF_2; 5.
DR PROSITE, PS01187; EGF_CA_2.
KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;
KW Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.
FT SIGNAL 1 28
FT CHAIN 29 1247 NIDOGEN.
FT DOMAIN 29 669 I (LARGER GLOBULAR DOMAIN).
FT DOMAIN 670 917 II (CYSTEINE-RICH).
FT DOMAIN 918 1247 III (SMALLER GLOBULAR DOMAIN).
FT DOMAIN 386 426 EGF-Like 1.
FT DOMAIN 668 709 EGF-Like 2.
FT DOMAIN 710 751 EGF-Like 3.
FT DOMAIN 758 801 EGF-Like 4.
FT DOMAIN 802 840 EGF-Like 5.
FT DOMAIN 872 919 CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 872 919 THYROGLOBULIN TYPE I.
FT DOMAIN 989 1030 LDL-RECEPTOR WYTD MOTIF 1.

```

Query Match	Match Local Similarity	27.18	Score 221	DB 1	Length 1247
Matches 54; Conservative 28; Mismatches 75; Indels 42; Gaps					
Qy	174	GPICITCDKMDGFFSLRNTHSICITACDSCKTSGLTRNDRGCECEVGMVLDEGACVYDDE	2333		
Db	658	GPV-----REGSPDALQNCITGTHACDTNNAACRGPRIQFCESIGFRGDRITCIDE	7133		
Qy	234	CAAEPPPCSAAPFCNANGSYTCBEC-----DSSCVCTGEGP-----GNCK--	2755		
Db	714	CSEGPSVCGSHITCINHHGTFRC-ECVGEYQFSDRGCTAVVDQRPINVCETGLHNCIDP	7727		
Qy	276	-----ECISGTAREHGGCADDVDECSLAETKTVKRNENCMYTPGSYVYVCP	3202		
Db	773	ORACQITVGGSSSYCTSCJPFSGSGGQACQDVDEC---QPSRCHPDACCTYNNPSFTQCK	8299		
Qy	321	DGFETEDACVPPAEAEAT	339		
Db	830	PEYQGDGFRCV-PGEVENT	847		
RESULT 32					
ID	NTCL_RAT	STANDARD:	PRT:	2531	AA.
AC	007008:				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Neurogenic locus notch homolog protein 1 precursor.				
GN	NOTCH1.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Schwann cell;				
RX	MEDLINE=92111383; PubMed=1764995;				
RA	Weinmaster G., Roberts V.J., Lemke G.;				
RT	"A homolog of Drosophila Notch expressed during mammalian				
RL	development".				
RN	Development 113:199-205(1991).				
RP	[2]				
REVISIONS TO 1652-1653.					


```

FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.

```

Query Match 11.08; Score 220; DB 1; Length 2531;

Best Local Similarity 23.6%; Pred. No. 8; 5e-09; Mismatches 97; Indels 126; Gaps 17;

```

Matches 76; Conservative 23;
QY 126 VC-CSPGTYPDCL---ACGGGSRP--CSGNHGCSGDSGRDSCRCRMGYGQPLC- 177
D 163 ICGCPHFHPTCRQDYNEC---SQNGILRHGCTCNE---ISYRCACRAHTGPHE 216
QY 178 -----TDCMDGYFSSLRNETHSICTA-----CDESKTCSGLTNRDCECEYGM 221
D 217 LPVYPCSPQCQNGGTCRPTGDTTNECACLRGFAGQNCCEWVDPCG-----NCKNG- 269
QY 222 VLDEGACY-----DYDECAERPPCSAAOFCKNANGSYTC----- 256
D 270 -----GACVGVNVTNCRCPPEMTGQYCTEDVDECOLMPNACQNAAGTCNHSHGYNVCYVN 325
QY 257 -----EEGDSGCVGCTGSG---PGNCKE-----CISGYAREHG 286
D 326 GMTGEDSCDNIIDPCASACFOGATCHDRVASFYCECPHGRTGLCHINDACTISNPNCEGS 385
QY 287 QC-----ADVDECSLAETKCVKRNENGYNTPGSYVCVCPDGF 323
D 386 NCDPNPVNGKAICTCPHGYTGPAQCSQDVDECALCANPCENHAK-CLMTLGSFECQCLGY 444
QY 324 -----EETEDACVP-PAEAENT 339
D 445 TGPRCEIDVNECISNCONDAT 466

```

```

RESULT 33
LML2_CABEL STANDARD; PRT; 3672 AA.
AC Q21313;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE laminin-like protein K08C7.3 precursor.
GN K08C7.3.
OS Caenorhabditis elegans.

```

```

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Berks M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1ab-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z70286; CA94293.1; -.
DR HSSP; P02468; IKLO.
DR WormPep; K08C7.3; CE06136.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00052; Laminin_B; 1.
DR Pfam; PF00053; Laminin_EGF; 21.
DR Pfam; PF00054; Laminin_G; 5.
DR Pfam; PF00055; Laminin_Nterm; 1.
DR PRINTS; PR00011; EGF/LAMININ.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 1.
DR SMART; SM00180; EGF_Lam; 21.
DR SMART; SM00281; LamG; 1.
DR SMART; SM00282; LamG; 5.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 21.
DR PROSITE; PS50025; LAM G DOMAIN; 5.
KW Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.
FT SIGNAL 1
FT CHAIN 27
FT 28 3672 POTENTIAL.
FT 28 297 LAMININ-LIKE PROTEIN K08C7.3.
FT DOMAIN 298 356 LAMININ N-TERMINAL (DOMAIN VI).
FT 357 426 LAMININ EGF-LIKE 1.
FT 427 471 LAMININ EGF-LIKE 2.
FT 472 477 LAMININ EGF-LIKE 3.
FT 478 518 LAMININ EGF-LIKE 4.
FT 519 563 LAMININ EGF-LIKE 5.
FT 564 609 LAMININ EGF-LIKE 6.
FT 610 655 LAMININ EGF-LIKE 7.
FT 656 700 LAMININ EGF-LIKE 8.
FT 701 755 LAMININ EGF-LIKE 9.
FT 756 808 LAMININ EGF-LIKE 10.
FT 809 839 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT 1415 1460 LAMININ EGF-LIKE 12.
FT 1461 1505 LAMININ EGF-LIKE 13.
FT 1506 1553 LAMININ EGF-LIKE 14.
FT 1554 1604 LAMININ EGF-LIKE 15.
FT 1605 1614 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT 1615 1796 LAMININ DOMAIN IV.
FT 1797 1829 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT 1830 1879 LAMININ EGF-LIKE 17.
FT 1880 1936 LAMININ EGF-LIKE 18.
FT 1937 1989 LAMININ EGF-LIKE 19.
FT 1990 2036 LAMININ EGF-LIKE 20.
FT 2037 2083 LAMININ EGF-LIKE 21.
FT 2084 2131 LAMININ EGF-LIKE 22.
FT 2693 2884 LAMININ G-LIKE 1.

```

FT	DOMAIN	2896	3066	LAMININ G-LIKE 2.
FT	DOMAIN	3072	3235	LAMININ G-LIKE 3.
FT	DOMAIN	3310	3482	LAMININ G-LIKE 4.
FT	DOMAIN	3488	3669	LAMININ G-LIKE 5.
FT	DISULFID	298	307	BY SIMILARITY.
FT	DISULFID	300	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	334	354	BY SIMILARITY.
FT	DISULFID	357	366	BY SIMILARITY.
FT	DISULFID	359	391	BY SIMILARITY.
FT	DISULFID	394	403	BY SIMILARITY.
FT	DISULFID	406	424	BY SIMILARITY.
FT	DISULFID	427	438	BY SIMILARITY.
FT	DISULFID	429	445	BY SIMILARITY.
FT	DISULFID	447	456	BY SIMILARITY.
FT	DISULFID	459	469	BY SIMILARITY.
FT	DISULFID	472	484	BY SIMILARITY.
FT	DISULFID	474	491	BY SIMILARITY.
FT	DISULFID	493	502	BY SIMILARITY.
FT	DISULFID	505	516	BY SIMILARITY.
FT	DISULFID	519	531	BY SIMILARITY.
FT	DISULFID	521	538	BY SIMILARITY.
FT	DISULFID	540	549	BY SIMILARITY.
FT	DISULFID	552	561	BY SIMILARITY.
FT	DISULFID	564	576	BY SIMILARITY.
FT	DISULFID	566	583	BY SIMILARITY.
FT	DISULFID	585	594	BY SIMILARITY.
FT	DISULFID	597	607	BY SIMILARITY.
FT	DISULFID	610	622	BY SIMILARITY.
FT	DISULFID	612	629	BY SIMILARITY.
FT	DISULFID	631	640	BY SIMILARITY.
FT	DISULFID	643	653	BY SIMILARITY.
FT	DISULFID	656	668	BY SIMILARITY.
FT	DISULFID	658	674	BY SIMILARITY.
FT	DISULFID	676	685	BY SIMILARITY.
FT	DISULFID	688	698	BY SIMILARITY.
FT	DISULFID	701	715	BY SIMILARITY.
FT	DISULFID	703	724	BY SIMILARITY.
FT	DISULFID	726	735	BY SIMILARITY.
FT	DISULFID	738	753	BY SIMILARITY.
FT	DISULFID	756	770	BY SIMILARITY.
FT	DISULFID	758	777	BY SIMILARITY.
FT	DISULFID	779	788	BY SIMILARITY.
FT	DISULFID	791	806	BY SIMILARITY.
FT	DISULFID	1415	1427	BY SIMILARITY.
FT	DISULFID	1417	1434	BY SIMILARITY.
FT	DISULFID	1436	1445	BY SIMILARITY.
FT	DISULFID	1448	1458	BY SIMILARITY.
FT	DISULFID	1461	1469	BY SIMILARITY.
FT	DISULFID	1463	1476	BY SIMILARITY.
FT	DISULFID	1478	1487	BY SIMILARITY.
FT	DISULFID	1490	1503	BY SIMILARITY.
FT	DISULFID	1506	1520	BY SIMILARITY.
FT	DISULFID	1508	1527	BY SIMILARITY.
FT	DISULFID	1529	1538	BY SIMILARITY.
FT	DISULFID	1541	1551	BY SIMILARITY.
FT	DISULFID	1554	1566	BY SIMILARITY.
FT	DISULFID	1575	1584	BY SIMILARITY.
FT	DISULFID	1587	1602	BY SIMILARITY.
FT	DISULFID	1830	1839	BY SIMILARITY.
FT	DISULFID	1832	1846	BY SIMILARITY.
FT	DISULFID	1849	1858	BY SIMILARITY.
FT	DISULFID	1861	1877	BY SIMILARITY.
FT	DISULFID	1880	1894	BY SIMILARITY.
FT	DISULFID	1882	1905	BY SIMILARITY.
FT	DISULFID	1907	1916	BY SIMILARITY.
FT	DISULFID	1919	1934	BY SIMILARITY.
FT	DISULFID	1937	1951	BY SIMILARITY.
FT	DISULFID	1939	1958	BY SIMILARITY.
FT	DISULFID	1961	1970	BY SIMILARITY.
FT	DISULFID	1973	1987	BY SIMILARITY.
FT	DISULFID	1990	2000	BY SIMILARITY.

FT	DISULFID	1992	2007	BY SIMILARITY.
FT	DISULFID	2009	2018	BY SIMILARITY.
FT	DISULFID	2021	2034	BY SIMILARITY.
FT	DISULFID	2037	2048	BY SIMILARITY.
FT	DISULFID	2039	2055	BY SIMILARITY.
FT	DISULFID	2057	2066	BY SIMILARITY.
FT	DISULFID	2069	2081	BY SIMILARITY.
FT	DISULFID	2084	2096	BY SIMILARITY.
FT	DISULFID	2103	2103	BY SIMILARITY.
FT	DISULFID	2105	2114	BY SIMILARITY.
FT	DISULFID	2117	2129	BY SIMILARITY.
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	140	140	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	249	249	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	351	351	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	477	477	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	511	511	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	634	634	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	761	761	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1014	1014	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1705	1705	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1756	1756	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1868	1868	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1944	1944	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1966	1966	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2002	2002	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2159	2159	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2207	2207	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2231	2231	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2235	2235	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2401	2401	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2421	2421	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2487	2487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2821	2821	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3087	3087	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3242	3242	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3541	3541	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	3672 AA;	404223 MW;	28E262DB5FF14BFA CRC64;
Query Match				
Best Local Similarity		23.18;	Score 217.5;	DB 1; Length 3672;
Matches		93; Conservative	31; Mismatches	170; Indels 109; Gaps 20;
OY	31	CHR-CRGLVDKRNQGVADRAKN-----FGGWTAW--EKKTLK-----YESS	71	
DB	331	CERCCPGFVQK--OMQAAIAHNNFTCEACNCFGRSNECEYDAEVLNKSIDSQGYEGG	388	
OY	72	EIRLLEILEGLCE----SSDEECNOMLEAEBHLEAW----LQKSEYPLDFEMFCV	121	
DB	389	-----GYCKNCRENTEGVNCKCSFYPRPEGVNTNEDPQCKVCCDDPKRHTGACA	439	
OY	122	KTLKVC-CSPGTYGPPLACQGG-----SORPCSGNHS-GDGSNQGSGSCRHNGYOG	174	
DB	440	EETGKCECPRFVGEDDCDCCASGYDAPCKPCECVNVTGIDVCLPEBGQCPCKAGFGG	499	
OY	175	PLCTDMDGFFSLRNETHSICACDESKCTSGLTNR-----DCGEVEGVWLD	224	
DB	500	TYCETACADGTWNTACVCEVCYCDATSGEHNCSASTGQCECKPATYAGLSCDKQCVGYFED	559	
OY	225	EGACVVDVDECAEPPPSAAQ--FCKMANGSYTCECD-----SSCVGCTGEGPG--	272	
DB	560	DCFKCNDPMGTGEGVCDQDTTGQCLCKBEPAGBKORCDAIARYGYNCAACADGAGTIS	619	
OY	273	-----NKECISGY-----AREHQCADVD-EC-----S	295	
DB	620	PECDATSGOCPCNGNFTGRTCDKCAAGFYNPDCRSCBCLLIGAKGQTCDSNGOCYCKGN	679	
OY	296	LAETGVARNENGYNPGSYVCV--PDGFETEDAC--PPAE	335	
DB	680	FBGERCDCKPKNFYNPICEBCNCNPSGVTRDFQGGCDKVSFGE	722	


```

DR PIR: A05267; A05267.
DR HSPB: P00740; 11XA.
DR FLYBASE: Fegm0004647; N.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_11.
DR InterPro: IPR000800; Notch.
DR Pfam: PF000023; ank; 6.
DR Pfam: PF000008; EGF; 36.
DR Pfam: PF000066; notch; 3.
DR PRINTS: PR001010; EGFBLDOD.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 4.
DR SMART: SM00179; EGF_Ca; 23.
DR SMART: SM00001; EGF_Like; 13.
DR SMART: SM00004; NL; 2.
DR PROSITE: PS50088; ANK_REPEAT; 5.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 22.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01187; EGF_Ca; 22.
DR PROSITE: PS01186; EGF_2; 28.
DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 44
FT CHAIN 45 2703
FT DOMAIN 45 1745
FT TRANSMEM 1746 1766
FT DOMAIN 1767 2703
FT DOMAIN 58 95
FT DOMAIN 96 136
FT DOMAIN 139 176
FT DOMAIN 177 215
FT DOMAIN 217 253
FT DOMAIN 255 291
FT DOMAIN 293 329
FT DOMAIN 331 370
FT DOMAIN 372 408
FT DOMAIN 409 447
FT DOMAIN 449 486
FT DOMAIN 488 524
FT DOMAIN 526 562
FT DOMAIN 564 600
FT DOMAIN 602 637
FT DOMAIN 639 675
FT DOMAIN 677 713
FT DOMAIN 715 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 865
FT DOMAIN 867 905
FT DOMAIN 907 944
FT DOMAIN 946 982
FT DOMAIN 984 1020
FT DOMAIN 1022 1058
Query Match 10.8%; Score 216.5; DB 1; Length 2703;
Best Local Similarity 23.3%; Pred. No. 1.7e-08;
Matches 97; Conservative 27; Mismatches 135; Indels 157; Gaps 20;
41 FNGGVDTAKK-----FGGNTAMEKTLTKY-----ESSEIR 74
130 FDSLSLEIAPNACDHVTLGNGTC--QLKTLLEYTCACANGYTGRCETKNLCASSPCR 187
75 LLEILGLCESSDFECN-----QMLEAGEHLEAWMLQKSEYDLEFEMFCVK 122
188 NGATCTALAGSSFTCSFGPGTGTCSYDIECQSN-----PCKYGGGCVW 224
123 T---LKVCCSPGTYPDC---LACGGGSGRPGSGNGHSGDGRGDSGCRHMGYGP 175

```

```

DR 235 THGSYQCMCPGTGYTKKPC----SPSPQNGIGCRNSGLST---ECKCPKGFBEK 288
QY 176 LCTDCMDGYFSSL-RNETHSICACDESKTCSGLTNRDCCGCEYGVWLDGACVYDDEC 234
DB 289 NGEQNTDCLGLHLCNGGCGICIDISDYTCRCPNFTGRCQD-----DYDEC 335
QY 235 A-AEPPCSAAOFCNKANGSYCEBCDSCV-GCTG---EPPGCKE-----CISG 280
DB 336 AQRDHPVCNGATCTNTHGSYSY----ICVNGWAGLDCSNNTDCKQAAFCYATCID 390
QY 281 VAREHGQCA-----IDV 291
DB 391 VGSFTQCTKTKTGLLCHLDADACTSPNCHADAICTDTPNNGSYACSCATGYGVCDSBI 450
QY 292 DECSIAEKTQVKNENCVTPGSSVYCVCPDGEETEDACVPAAEATGEGS-PTQ 346
DB 451 DECD--QGSPPCEHNGICVMTPOSYNRNCSGQFTG-----PCEFNINCESHPQ 498
RESULT 35
FBI5 RAT STANDARD; PRT; 448 AA.
ID FBI5 RAT
AC 09WVH8: 09R284;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Ftblin-5 precursor (Ftbl-5) (Developmental arteries and neural crest
DE EGF-like protein) (Dance) (Embryonic vascular EGF repeat-containing
DE protein) (EVEC).
GN FBI5 OR DANCE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid:10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:99357779; PubMed:10428823;
RA Nakamura T., Ruiz-Lozano P., Lindner V., Yabe D., Tanikawa M.,
RA Furukawa Y., Kobuke K., Tashiro K., Lu Z., Andon N.L., Schaub R.,
RA Matsumori A., Sasayama S., Chien K.R., Honjo T.;
RT "DANCE, a novel secreted RGD protein expressed in developing,
RT atherosclerotic, and balloon-injured arteries.";
RT J. Biol. Chem. 274:22476-22483(1999).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE:99278197; PubMed:10347091;
RA Kowal R.C., Richardson J.A., Milano J.M., Olson E.N.;
RT "EVEC, a novel epidermal growth factor-like repeat-containing protein
RT upregulated in embryonic and diseased adult vasculature.";
RT Circ. Res. 84:1166-1176(1999).
RL -!- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
INTERACTION OF INTEGRINS AND THE RGD MOTIF. COULD BE A VASCULAR
LIGAND FOR INTEGRIN RECEPTORS AND MAY PLAY A ROLE IN VASCULAR
DEVELOPMENT AND REMODELING.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL: AF12153; AAD41769.1; -
DB EMBL: AF137350; AAD25101.1; -
DB HSSP: P00736; IADQ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF_Like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00008; EGF; 4.
DR SMART: SM00179; EGF_Ca; 4.

```

DR SMART; SM00001; EGF_Like; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS01187; EGF_CA; 6.
 KW Repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 448
 FT DOMAIN 24 69
 FT DOMAIN 127 167
 FT DOMAIN 168 206
 FT DOMAIN 207 246
 FT DOMAIN 247 287
 FT DOMAIN 288 333
 FT SITE 54 56
 FT DISULFID 131 144
 FT DISULFID 138 153
 FT DISULFID 155 166
 FT DISULFID 172 181
 FT DISULFID 177 190
 FT DISULFID 192 205
 FT DISULFID 211 221
 FT DISULFID 217 230
 FT DISULFID 232 245
 FT DISULFID 251 262
 FT DISULFID 258 271
 FT DISULFID 273 286
 FT DISULFID 292 305
 FT DISULFID 299 314
 FT DISULFID 320 332
 FT CARBOHYD 283 283
 FT CARBOHYD 296 296
 FT CONFLICT 238 238
 SQ SEQUENCE 448 AA; 50160 MM; E6BC68F7B14B714 CRC64;
 L -> P (IN REF. 2). (POTENTIAL).

Query Match 10.8%; Score 216; DB 1; Length 448;
 Best Local Similarity 29.3%; Pred. No. 3,4e-09;
 Matches 54; Conservative 12; Mismatches 52; Indels 66; Gaps 8;
 QY 217 CEVGMVLDG-ACVDVDECAEAPPSCAOPCKNANGSYTC----- 257
 DB 113 CRFGYQMDGNCQVDECAIDSHQCNPQICINTEGGYTCSDGYWLLGQCLIDDEC 172
 QY 258 ---ECDSVCYCGTGE-----GPG-----NCKE-----C 277
 DB 173 RYGYCQQLCANVPSSYSCTCNPFTLNDGRSCQDVNECETENPCVOTCVMTYGSFTICRC 232
 QY 278 ISGYAREHG--QCADVDECSLAERTCYRKNKNCYNTPGSYVCPPDGEETED--ACVPP 333
 DB 233 DPGYLEEDGTHCSMDMECSFELC---QHECVNPGSFTSCPPGVLLLEDNRSQDI 289
 QY 334 AEAE 337
 DB 290 NECE 293
 RESULT 36
 FBL5_MOUSE
 ID FBL5_MOUSE STANDARD; PRT; 448 AA.
 AC Q9WH9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE FIBULIN-5 precursor (FIBL-5) (Developmental arteries and neural crest
 DE EGF-like protein) (Dance).
 GN FBLN5 OR DANCE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=99357779; PubMed=10428823;
 RA Nakamura T., Ruiz-Lozano P., Lindner V., Yabe D., Taniwaki M.,
 RA Furukawa Y., Kobuke K., Tashiro K., Lu Z., Andon N.L., Schaub R.,
 RA Matsumori A., Sasayama S., Chien K.R., Honjo T.;
 RT "DANCE, a novel secreted RGD protein expressed in developing,
 RT atherosclerotic, and balloon-injured arteries."
 RL J. Biol. Chem. 274:22476-22483(1999).
 CC -1- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
 CC INTERACTION OF INTEGRINS AND THE RGD MOTIF. COULD BE A VASCULAR
 CC LIGAND FOR INTEGRIN RECEPTORS AND MAY PLAY A ROLE IN VASCULAR
 CC DEVELOPMENT AND REMODELING.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF112151; AAD41767.1; -;
 DR HSSP; P00736; IAPQ.
 DR MGD; MG1:1346091; Fbln5.
 DR Interpro; IPR000152; Asx_hydroxyl.
 DR Interpro; IPR000561; EGF-like.
 DR Interpro; IPR001881; EGF_Ca.
 DR Pfam; PF00008; EGF; 4.
 DR SMART; SM00179; EGF_CA; 4.
 DR SMART; SM00001; EGF_Like; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS01187; EGF_CA; 6.
 KW Repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 448
 FT DOMAIN 24 69
 FT DOMAIN 127 167
 FT DOMAIN 168 206
 FT DOMAIN 207 246
 FT DOMAIN 247 287
 FT DOMAIN 288 333
 FT SITE 54 56
 FT DISULFID 131 144
 FT DISULFID 138 153
 FT DISULFID 155 166
 FT DISULFID 172 181
 FT DISULFID 177 190
 FT DISULFID 192 205
 FT DISULFID 211 221
 FT DISULFID 217 230
 FT DISULFID 232 245
 FT DISULFID 251 262
 FT DISULFID 258 271
 FT DISULFID 273 286
 FT DISULFID 292 305
 FT DISULFID 299 314
 FT DISULFID 320 332
 FT CARBOHYD 283 283
 FT CARBOHYD 296 296
 SQ SEQUENCE 448 AA; 50193 MM; P15CC70CCFBDC97 CRC64;
 L -> P (IN REF. 2). (POTENTIAL).

Query Match 10.7%; Score 214.5; DB 1; Length 448;
 Best Local Similarity 36.2%; Pred. No. 4,3e-09;
 Matches 47; Conservative 7; Mismatches 37; Indels 39; Gaps 5;
 QY 217 CEVGMVLDG-ACVDVDECAEAPPSCAOPCKNANGSYTCBECDSVCGTGSGPNCK 275
 DB 113 CRFGYQMDGNCQVDECAIDSHQCNPQICINTEGGYTC----- 153

OY 276 ECISGAREHGOCADVDECSIAEKTCKRNCNCYTPGSGVCPDGEF----- 324
 1 11 111 111 1 111 111 11
 DB 154 SCTDGVWLLGGCCLDIDECRYG--YC---QQLCANVPGSYCTCNGNGLINDGRSCQDV 208
 OY 325 ---ETEDACV 331
 111: 11
 DB 209 NECETENPCV 218
 RESULT 37
 NIDO_MOUSE STANDARD: PRT: 1245 AA.
 AC P10493:
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nidogen precursor (Entactin).
 GN NID OR NID1 OR ENT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40.
 RX MEDLINE=89079780; PubMed=3264556;
 RA Durkin M.E., Chakravarti S., Bartos B.B., Liu S.H., Friedman R.L.,
 RA Chung A.E.;
 RT "Amino acid sequence and domain structure of entactin. Homology with
 RT epidermal growth factor precursor and low density lipoprotein
 RT receptor.";
 RL J. Cell Biol. 107:2749-2756(1988).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89231638; PubMed=2496973;
 RA Mann K., Deutzmann R., Aumailley M., Timpl R., Raimondi L.,
 RA Yamada Y., Pan T.-C., Conway D., Chu M.-L.;
 RT "Amino acid sequence of mouse nidogen, a multidomain basement
 RT membrane protein with binding activity for laminin, collagen IV and
 RT cells.";
 RL EMBO J. 8:65-72(1989).
 RN [3]
 RP SEQUENCE OF 1-251 FROM N.A.
 RC STRAIN=BAUB/C; TISSUE=Liver;
 RX MEDLINE=94040771; PubMed=8224873;
 RA Durkin M.E., Liu S.H., Reing J., Chung A.E.;
 RT "Characterization of the 5' end of the mouse Ent gene encoding the
 RT basement membrane protein, entactin.";
 RL Gene 132:261-266(1993).
 RN [4]
 RP SEQUENCE OF 1207-1245 FROM N.A.
 RC STRAIN=C57BL/6J X CBA/J;
 RX MEDLINE=95324912; PubMed=7601446;
 RA Durkin M.E., Wewer U.M., Chung A.E.;
 RT "Exon organization of the mouse entactin gene corresponds to the
 RT structural domains of the polypeptide and has regional homology to
 RT the low-density lipoprotein receptor gene.";
 RL Genomics 26:219-228(1995).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=86192477; PubMed=3084254;
 RA Paulsson M., Deutzmann R., Dziadek M., Nowack H., Timpl R., Weber S.,
 RA Engel J.;
 RT "Purification and structural characterization of intact and
 RT fragmented nidogen obtained from a tumor basement membrane.";
 RL Eur. J. Biochem. 156:467-478(1986).
 RN [6]
 RP CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.
 RX MEDLINE=93316903; PubMed=8326911;
 RA Fujiwara S., Shinkai H., Mann K., Timpl R.;
 RT "Structure and localization of O- and N-linked oligosaccharide chains
 RT on basement membrane protein nidogen.";
 RL Matrix 13:215-222(1993).
 CG -1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN

CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.
 CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-
 CC EXTRACELLULAR MATRIX INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
 CC -1- PTM: N- AND O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-1 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X14194; CAA32408.1; -;
 DR EMBL: X14480; CAA32642.1; -;
 DR EMBL: L17324; AAB7652.1; -;
 DR EMBL: L17322; AAB7652.1; JOINED.
 DR EMBL: L17323; AAB7652.1; JOINED.
 DR EMBL: X83093; CAA58148.1; -;
 DR PIR: S02730; MAMMSND.
 DR HSSP: P07204; IADX.
 DR CarbBank; CCSD:37018; -;
 DR CarbBank; CCSD:37019; -;
 DR MGD: MGI:973442; Nid1.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR000033; ldl_receptor_rep.
 DR InterPro: IPR003886; Nidogen_ext.
 DR InterPro: IPR000716; Thyroglobulin_1.
 DR Pfam: PF00008; EGF_6.
 DR Pfam: PF00058; ldl_recept_b; 3.
 DR Pfam: PF00086; thyroglobulin_1; 1.
 DR SMART: SM00179; EGF_CA; 2.
 DR SMART: SM00001; EGF_Like; 4.
 DR SMART: SM00135; LY; 5.
 DR SMART: SM00539; NIDO; 1.
 DR SMART: SM00211; TY; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 3.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
 DR PROSITE: PS01186; EGF_2; 4.
 DR PROSITE: PS01187; EGF_CA; 2.
 KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;
 KW Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.
 KM SIGNAL 1 28
 FT CHAIN 29 1245
 FT DOMAIN 29 667 I (LARGER GLOBULAR DOMAIN).
 FT DOMAIN 668 919 II (CYSTEINE-RICH).
 FT DOMAIN 920 1245 III (SMALLER GLOBULAR DOMAIN).
 FT DOMAIN 384 424 EGF-LIKE 1.
 FT DOMAIN 666 707 EGF-LIKE 2.
 FT DOMAIN 708 749 EGF-LIKE 3. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 756 799 EGF-LIKE 4.
 FT DOMAIN 800 838 EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 870 917 THYROGLOBULIN TYPE 1.
 FT DOMAIN 967 1028 LDL-RECEPTOR YWTD MOTIF 1.
 FT DOMAIN 1030 1071 LDL-RECEPTOR YWTD MOTIF 2.
 FT DOMAIN 1073 1116 LDL-RECEPTOR YWTD MOTIF 3.
 FT DOMAIN 1122 1161 LDL-RECEPTOR YWTD MOTIF 4.
 FT DOMAIN 1206 1242 EGF-LIKE 6.
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 290 290 SULFATION (POTENTIAL).
 FT MOD_RES 295 295 SULFATION (POTENTIAL).
 FT DISULFD 670 683 BY SIMILARITY.
 FT DISULFD 677 693 BY SIMILARITY.
 FT DISULFD 695 706 BY SIMILARITY.
 FT DISULFD 712 725 BY SIMILARITY.

FT	DOMAIN	1123	1159	EGF-LIKE 29.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1161	1204	EGF-LIKE 30.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1206	1245	EGF-LIKE 31.	
FT	DOMAIN	1247	1288	EGF-LIKE 32.	
FT	DOMAIN	1290	1326	EGF-LIKE 33.	
FT	DOMAIN	1336	1374	EGF-LIKE 34.	
FT	REPEAT	1388	1428	LIN/NOTCH 1.	
FT	REPEAT	1429	1467	LIN/NOTCH 2.	
FT	REPEAT	1468	1503	LIN/NOTCH 3.	
FT	REPEAT	1839	1868	ANK 1.	
FT	REPEAT	1872	1902	ANK 2.	
FT	REPEAT	1906	1935	ANK 3.	
FT	REPEAT	1939	1968	ANK 4.	
FT	REPEAT	1972	2001	ANK 5.	
FT	DISULFID	43	55	BY SIMILARITY.	
FT	DISULFID	49	66	BY SIMILARITY.	
FT	DISULFID	68	77	BY SIMILARITY.	
FT	DISULFID	83	94	BY SIMILARITY.	
FT	DISULFID	88	107	BY SIMILARITY.	
FT	DISULFID	109	118	BY SIMILARITY.	
FT	DISULFID	124	135	BY SIMILARITY.	
FT	DISULFID	129	145	BY SIMILARITY.	
FT	DISULFID	147	156	BY SIMILARITY.	
FT	DISULFID	163	175	BY SIMILARITY.	
FT	DISULFID	169	184	BY SIMILARITY.	
FT	DISULFID	186	195	BY SIMILARITY.	
FT	DISULFID	202	213	BY SIMILARITY.	
FT	DISULFID	207	223	BY SIMILARITY.	
FT	DISULFID	225	234	BY SIMILARITY.	
FT	DISULFID	241	252	BY SIMILARITY.	
FT	DISULFID	246	261	BY SIMILARITY.	
FT	DISULFID	263	272	BY SIMILARITY.	
FT	DISULFID	279	292	BY SIMILARITY.	
FT	DISULFID	286	301	BY SIMILARITY.	
FT	DISULFID	303	312	BY SIMILARITY.	
FT	DISULFID	319	330	BY SIMILARITY.	
FT	DISULFID	324	339	BY SIMILARITY.	
FT	DISULFID	341	350	BY SIMILARITY.	
FT	DISULFID	356	367	BY SIMILARITY.	
FT	DISULFID	361	378	BY SIMILARITY.	
FT	DISULFID	380	389	BY SIMILARITY.	
FT	DISULFID	396	409	BY SIMILARITY.	
FT	DISULFID	403	418	BY SIMILARITY.	
FT	DISULFID	420	429	BY SIMILARITY.	
FT	DISULFID	436	447	BY SIMILARITY.	
FT	DISULFID	441	456	BY SIMILARITY.	
FT	DISULFID	458	467	BY SIMILARITY.	
FT	DISULFID	474	485	BY SIMILARITY.	
FT	DISULFID	479	494	BY SIMILARITY.	
FT	DISULFID	496	505	BY SIMILARITY.	
FT	DISULFID	512	523	BY SIMILARITY.	
FT	DISULFID	517	532	BY SIMILARITY.	
FT	DISULFID	534	543	BY SIMILARITY.	
FT	DISULFID	550	560	BY SIMILARITY.	
FT	DISULFID	555	569	BY SIMILARITY.	
FT	DISULFID	571	580	BY SIMILARITY.	
FT	DISULFID	587	598	BY SIMILARITY.	
FT	DISULFID	592	607	BY SIMILARITY.	
FT	DISULFID	609	618	BY SIMILARITY.	
FT	DISULFID	625	635	BY SIMILARITY.	
FT	DISULFID	630	644	BY SIMILARITY.	
FT	DISULFID	646	655	BY SIMILARITY.	
FT	DISULFID	662	673	BY SIMILARITY.	
FT	DISULFID	667	682	BY SIMILARITY.	
FT	DISULFID	684	693	BY SIMILARITY.	
FT	DISULFID	700	710	BY SIMILARITY.	
FT	DISULFID	705	719	BY SIMILARITY.	
FT	DISULFID	721	730	BY SIMILARITY.	
FT	DISULFID	739	750	BY SIMILARITY.	
FT	DISULFID	744	759	BY SIMILARITY.	
FT	DISULFID	761	770	BY SIMILARITY.	
FT	DISULFID	776	787	BY SIMILARITY.	
FT	DISULFID	781	797	BY SIMILARITY.	

Query Match	Best Local	Similarity	10.6%	Score 212;	DB 1;	Length 2318;
Matches 73;	Conservative	30;	Mismatches 86;	Indels 70;	Gaps	
QY 128 CSPGTYGPDCL---	ACQSGSORPCSGNGHCSGD---	GSNGSGSGCRCHMGYGPCLTD--	179			
Db 68 CLIPGVWGRCOLEDPCSHG--	PCAGRGVCOSSVVAQTAR--	FSCRCILRGFGQPDSSQPD	122			
QY 180 -----CNDGXFSSLRNETHSICTAC-----	DESCDT-----C-SGLTNRDCG-----	215				
Db 123 PCVSRPCYCHGAPCSGPPDRFAC-ACPRGYQSQSCSDIDERSGTYCHGGTCLNTPS	181					
QY 216 ---ECEVGM--VLDEGACVDVDECAAEPPPCSAOFCNKANG--SYTC-----	EECDSS	261				
Db 182 FRCCGFLATGYGLTCENPVY---PCA--PSPCRNNGGTCRQSSDPTYDCACTLPREGONCEY	236					
QY 262 SCVCGTGGSPGNCCKCTISGYAREHQC-----	ADVDECSLAETKCYKKNENCYNT	311				
Db 237 NVDDCPGHRCLNGGTCYGVGVNTNYNCOCPPEWTOGCTEDVDYDECQIDPNAC-HNGGTCFNL	295					
QY 312 PGSYVVCVPDPEFETEDAC	330					
Db 296 LGGHSCVANGW--TGESC	312					
RESULT 39						
NEEL_HUMAN						
AC Q92832; Q9Y472;	STANDARD;	PRT;	810 AA.			
DT 01-NOV-1997 (Rel. 35, Created)						
DT 16-OCT-2001 (Rel. 40, Last sequence update)						
DT 16-OCT-2001 (Rel. 40, Last annotation update)						
DE Protein kinase C-binding protein NEEL1 precursor (NEEL-like protein 1)						
DE (Nei-related protein 1).						
GN NEEL1 OR NRPL						
OS Homo sapiens (Human).						
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
OX NCBI_TaxID=9606;						
NP [1]						
SEQUENCE FROM N.A.						

CC TISSUE-Brain;
RX MEDLINE-97131504; PubMed-8975702;
RA Matanabe T.K., Katagiri T., Suzuki M., Shimizu F., Fujiwara T.,
RA Kanemoto N., Nakamura Y., Hirai Y., Maekawa H., Takahashi E.,
RT "Cloning and characterization of two novel human CDNs (NELL1 and
NELL2) encoding proteins with six EGF-like repeats."
RL Genomics 38:273-276(1996).
RN [2]
RP SEQUENCE OF 383-810 FROM N.A.
RA Ting K., Vastardis H., Mulliken J.B., Bertolami C., Wen Z.,
RA Young M., Tleu A., Kwong E.,
RT "Nel homolog gene expression in craniofacial anomalies."
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1 SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1 (BY SIMILARITY).
CC -1 SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1 (BY SIMILARITY).
CC -1 SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1 (BY SIMILARITY).
CC -1 DISEASE: EXPRESSED IN CRANIOFACIAL ANOMALIES.
CC -1 SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
CC -1 SIMILARITY: CONTAINS 5 WMFC DOMAINS.
CC -1 SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -1 CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
IN POSITIONS 427 AND 771.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D83017; BAAL1680.1; .
DR EMBL; U57523; AAB06946.1; ALT_FRAME.
DR HSSP; P07204; IADX.
DR MIM; 602319; .
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WMFC.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00093; vwc; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_Like; 4.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; vwc; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01208; WMFC; 2.
KW Glycoprotein; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 16
FT CHAIN 17 810
FT DOMAIN 81 230
FT DOMAIN 273 331
FT DOMAIN 335 390
FT DOMAIN 391 433
FT DOMAIN 434 475
FT DOMAIN 476 516
FT DOMAIN 515 547
FT DOMAIN 549 595
FT DOMAIN 596 631
FT DOMAIN 632 687
FT DOMAIN 687 750
FT DOMAIN 752 807
FT DOMAIN 807 847
FT DISULFID 395 407
FT DISULFID 401 416
FT DISULFID 418 432
FT DISULFID 438 451
FT DISULFID 445 460
FT DISULFID 462 474

FT DISULFID 480 493 BY SIMILARITY.
FT DISULFID 487 502 BY SIMILARITY.
FT DISULFID 504 515 BY SIMILARITY.
FT DISULFID 519 529 BY SIMILARITY.
FT DISULFID 523 535 BY SIMILARITY.
FT DISULFID 537 546 BY SIMILARITY.
FT DISULFID 553 566 BY SIMILARITY.
FT DISULFID 560 575 BY SIMILARITY.
FT DISULFID 577 594 BY SIMILARITY.
FT DISULFID 600 613 BY SIMILARITY.
FT DISULFID 607 622 BY SIMILARITY.
FT DISULFID 624 630 BY SIMILARITY.
FT DISULFID 640 640 BY SIMILARITY.
FT CARBOHYD 53 53 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 83 83 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 224 224 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 562 562 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 609 609 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 732 732 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 758 758 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CONFLICT 383 383 N -> D (IN REF. 2).
FT CONFLICT 573 573 Y -> H (IN REF. 2).
FT CONFLICT 626 626 S -> C (IN REF. 2).
SQ SEQUENCE 810 AA; 89606 MM; 549465EASJF7AEED0 CRC64;

Query Match 10.5%; Score 211.5; DB 1; Length 810;
Best Local Similarity 22.0%; Pred. No. 1.3e-08;
Matches 95; Conservative 30; Mismatches 122; Indels 185; Gaps 21;

OY 28 PTPCHRCRLVDKFN--QGMVD-----TAKNNGGNTAMEKTSKYSSSEIRLE 77
DB 221 PNLNHTCPTCSDFLSLVQGIIMDLAKMTAKLYVA-----ETRLSQLENCHEKTC 273
OY 78 ILEGL-----CSDSDEFCNOMLEAGEHLEAMWLQKSEYPLDFEMF 119
DB 274 QVSGLLYRDQDSWDGDHCRNCTCKSGAVECRMM----- 307
OY 120 CVKTLKVCSPCTGYPRDCLACGGGSGQ-----NP-CSGNHCSGDSRQDSCRCRMGYQ 173
DB 308 -----SCPLNCSPPDLSPIVIAAGCCVCKPCITGYGKVLAEORILTKSCR----- 354
OY 174 GPLCTDCMDGYSSLRNETHSI-CTACD-----ESKTCGSLT-----NRDGS- 215
DB 355 -----ECRGGVLVKITEMCPRPLNCSEKDIILPENOCRCVRGRHNFCAEGPKGENSECKN 409
OY 216 -----RCVEGVWVL-----DEGACVVDDECAARPPCSAOPFCNNNGSYTCECCSSCV- 264
DB 410 WNTFKATCECKSGSYISVQGSAYCEDIDECBAAMHYCHANTVAVNLPGLYRCD-----CVP 464
OY 265 -----GCT-----GEGPGNCKE-----CISGY-----AR 283
DB 465 GYIRVDPSCPTHDECGSQHNCNDENALCTNTVOGHSCRCRGYGNCTICARFCEEGCR 524
OY 284 EHGGCA-----DVDECSLAKEKTCVRKNENCYNTPGSGYCVCPDGEFE 325
DB 525 YGCTVAPARKVCVPSGFTGSHCEKIDIDECSEGIIEC-HNHSRCVNLPGWYHCECHRSRGFHD 583
OY 326 -----TEDACV 331
DB 584 DGTYSLSGESCT 595

RESULT 40
EMBL_MOUSE STANDARD; PRT; 931 AA.
ID EMBL_MOUSE
AC 061549;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)

DT 01-MAR-2002 (Rel. 41, last annotation update)
DE Cell surface glycoprotein EMRI precursor (EMRI hormone receptor)
GN EMRI OR GPF480.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C; TISSUE=Peritoneal cavity;
RX MEDLINE=96132946; PubMed=8350607;
RA McKnight A.J., Macfarlane A.J., Dri P., Turley L., Willis A.C.,
RA Gordon S.;
RT "Molecular cloning of F4/80, a murine macrophage-restricted cell
RT surface glycoprotein with homology to the G-protein-linked
RT transmembrane 7 hormone receptor family.";
RL J. Biol. Chem. 271:486-489(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97312684; PubMed=9169125;
RA Lin H.H., Stubbs L.J., Mucenski M.L.;
RT "Identification and characterization of a seven transmembrane hormone
RT receptor using differential display.";
RL Genomics 41:301-308(1997).
CC -1- FUNCTION: PROBABLY INVOLVED IN CELL ADHESION WITHIN TISSUES
CC AND RECEPTOR SIGNALING.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: IN MACROPHAGES, BUT ABSENT FROM THOSE WHICH
CC ARE LOCALIZED WITHIN T-CELL AREAS OF LYMPH NODES AND SPLEEN.
CC LOW LEVEL OF EXPRESSION ON BLOOD MONOCYTES.
CC -1- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL: X93328; CAA63720.1; -
DR EMBL: U66888; AAC53184.1; -
DR HSSP: P07204; 1FGD.
DR GCRDB: GCR_1309; -
DR MGD: MGI:106912; Emr1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR000832; GPCR_secretin.
DR InterPro: IPR000203; PKD_cys-rich.
DR Pfam: PF00002; 7tm_2; 1.
DR Pfam: PF00008; EGF; 6.
DR Pfam: PF01825; GPS; 1.
DR PRINTS: PRO1128; EMRIHORMONER.
DR SMART: SM00179; EGF_CA; 6.
DR SMART: SM00001; EGF_Like; 1.
DR SMART: SM00303; GPS; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 6.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 5.
DR PROSITE: PS00221; GPS; 1.
DR PROSITE: PS00650; G_PROTEIN_RECP_F2_4; 1.
DR PROSITE: PS50261; G_PROTEIN_RECP_F2_4; 1.
DR G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;
KW EGF-like domain; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 931
FT DOMAIN 28 644
FT TRANSMEM 645 672
FT DOMAIN 673 679
FT POTENTIAL.
FT CELL SURFACE GLYCOPROTEIN EMRI.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 680 701
FT DOMAIN 702 711
FT TRANSMEM 712 735
FT DOMAIN 736 754
FT TRANSMEM 755 776
FT DOMAIN 777 792
FT TRANSMEM 793 821
FT DOMAIN 822 839
FT TRANSMEM 840 859
FT DOMAIN 860 874
FT TRANSMEM 875 897
FT DOMAIN 898 931
FT TRANSMEM 932 954
FT DOMAIN 955 976
FT TRANSMEM 977 998
FT DOMAIN 999 1020
FT TRANSMEM 1021 1042
FT DOMAIN 1043 1064
FT TRANSMEM 1065 1086
FT DOMAIN 1087 1108
FT TRANSMEM 1109 1130
FT DOMAIN 1131 1152
FT TRANSMEM 1153 1174
FT DOMAIN 1175 1196
FT TRANSMEM 1197 1218
FT DOMAIN 1219 1240
FT TRANSMEM 1241 1262
FT DOMAIN 1263 1284
FT TRANSMEM 1285 1306
FT DOMAIN 1307 1328
FT TRANSMEM 1329 1350
FT DOMAIN 1351 1372
FT TRANSMEM 1373 1394
FT DOMAIN 1395 1416
FT TRANSMEM 1417 1438
FT DOMAIN 1439 1460
FT TRANSMEM 1461 1482
FT DOMAIN 1483 1504
FT TRANSMEM 1505 1526
FT DOMAIN 1527 1548
FT TRANSMEM 1549 1570
FT DOMAIN 1571 1592
FT TRANSMEM 1593 1614
FT DOMAIN 1615 1636
FT TRANSMEM 1637 1658
FT DOMAIN 1659 1680
FT TRANSMEM 1681 1702
FT DOMAIN 1703 1724
FT TRANSMEM 1725 1746
FT DOMAIN 1747 1768
FT TRANSMEM 1769 1790
FT DOMAIN 1791 1812
FT TRANSMEM 1813 1834
FT DOMAIN 1835 1856
FT TRANSMEM 1857 1878
FT DOMAIN 1879 1900
FT TRANSMEM 1901 1922
FT DOMAIN 1923 1944
FT TRANSMEM 1945 1966
FT DOMAIN 1967 1988
FT TRANSMEM 1989 2010
FT DOMAIN 2011 2032
FT TRANSMEM 2033 2054
FT DOMAIN 2055 2076
FT TRANSMEM 2077 2098
FT DOMAIN 2099 2120
FT TRANSMEM 2121 2142
FT DOMAIN 2143 2164
FT TRANSMEM 2165 2186
FT DOMAIN 2187 2208
FT TRANSMEM 2209 2230
FT DOMAIN 2231 2252
FT TRANSMEM 2253 2274
FT DOMAIN 2275 2296
FT TRANSMEM 2297 2318
FT DOMAIN 2319 2340
FT TRANSMEM 2341 2362
FT DOMAIN 2363 2384
FT TRANSMEM 2385 2406
FT DOMAIN 2407 2428
FT TRANSMEM 2429 2450
FT DOMAIN 2451 2472
FT TRANSMEM 2473 2494
FT DOMAIN 2495 2516
FT TRANSMEM 2517 2538
FT DOMAIN 2539 2560
FT TRANSMEM 2561 2582
FT DOMAIN 2583 2604
FT TRANSMEM 2605 2626
FT DOMAIN 2627 2648
FT TRANSMEM 2649 2670
FT DOMAIN 2671 2692
FT TRANSMEM 2693 2714
FT DOMAIN 2715 2736
FT TRANSMEM 2737 2758
FT DOMAIN 2759 2780
FT TRANSMEM 2781 2802
FT DOMAIN 2803 2824
FT TRANSMEM 2825 2846
FT DOMAIN 2847 2868
FT TRANSMEM 2869 2890
FT DOMAIN 2891 2912
FT TRANSMEM 2913 2934
FT DOMAIN 2935 2956
FT TRANSMEM 2957 2978
FT DOMAIN 2979 3000
FT TRANSMEM 3001 3022
FT DOMAIN 3023 3044
FT TRANSMEM 3045 3066
FT DOMAIN 3067 3088
FT TRANSMEM 3089 3110
FT DOMAIN 3111 3132
FT TRANSMEM 3133 3154
FT DOMAIN 3155 3176
FT TRANSMEM 3177 3198
FT DOMAIN 3199 3220
FT TRANSMEM 3221 3242
FT DOMAIN 3243 3264
FT TRANSMEM 3265 3286
FT DOMAIN 3287 3308
FT TRANSMEM 3309 3330
FT DOMAIN 3331 3352
FT TRANSMEM 3353 3374
FT DOMAIN 3375 3396
FT TRANSMEM 3397 3418
FT DOMAIN 3419 3440
FT TRANSMEM 3441 3462
FT DOMAIN 3463 3484
FT TRANSMEM 3485 3506
FT DOMAIN 3507 3528
FT TRANSMEM 3529 3550
FT DOMAIN 3551 3572
FT TRANSMEM 3573 3594
FT DOMAIN 3595 3616
FT TRANSMEM 3617 3638
FT DOMAIN 3639 3660
FT TRANSMEM 3661 3682
FT DOMAIN 3683 3704
FT TRANSMEM 3705 3726
FT DOMAIN 3727 3748
FT TRANSMEM 3749 3770
FT DOMAIN 3771 3792
FT TRANSMEM 3793 3814
FT DOMAIN 3815 3836
FT TRANSMEM 3837 3858
FT DOMAIN 3859 3880
FT TRANSMEM 3881 3902
FT DOMAIN 3903 3924
FT TRANSMEM 3925 3946
FT DOMAIN 3947 3968
FT TRANSMEM 3969 3990
FT DOMAIN 3991 4012
FT TRANSMEM 4013 4034
FT DOMAIN 4035 4056
FT TRANSMEM 4057 4078
FT DOMAIN 4079 4100
FT TRANSMEM 4101 4122
FT DOMAIN 4123 4144
FT TRANSMEM 4145 4166
FT DOMAIN 4167 4188
FT TRANSMEM 4189 4210
FT DOMAIN 4211 4232
FT TRANSMEM 4233 4254
FT DOMAIN 4255 4276
FT TRANSMEM 4277 4298
FT DOMAIN 4299 4320
FT TRANSMEM 4321 4342
FT DOMAIN 4343 4364
FT TRANSMEM 4365 4386
FT DOMAIN 4387 4408
FT TRANSMEM 4409 4430
FT DOMAIN 4431 4452
FT TRANSMEM 4453 4474
FT DOMAIN 4475 4496
FT TRANSMEM 4497 4518
FT DOMAIN 4519 4540
FT TRANSMEM 4541 4562
FT DOMAIN 4563 4584
FT TRANSMEM 4585 4606
FT DOMAIN 4607 4628
FT TRANSMEM 4629 4650
FT DOMAIN 4651 4672
FT TRANSMEM 4673 4694
FT DOMAIN 4695 4716
FT TRANSMEM 4717 4738
FT DOMAIN 4739 4760
FT TRANSMEM 4761 4782
FT DOMAIN 4783 4804
FT TRANSMEM 4805 4826
FT DOMAIN 4827 4848
FT TRANSMEM 4849 4870
FT DOMAIN 4871 4892
FT TRANSMEM 4893 4914
FT DOMAIN 4915 4936
FT TRANSMEM 4937 4958
FT DOMAIN 4959 4980
FT TRANSMEM 4981 5002
FT DOMAIN 5003 5024
FT TRANSMEM 5025 5046
FT DOMAIN 5047 5068
FT TRANSMEM 5069 5090
FT DOMAIN 5091 5112
FT TRANSMEM 5113 5134
FT DOMAIN 5135 5156
FT TRANSMEM 5157 5178
FT DOMAIN 5179 5200
FT TRANSMEM 5201 5222
FT DOMAIN 5223 5244
FT TRANSMEM 5245 5266
FT DOMAIN 5267 5288
FT TRANSMEM 5289 5310
FT DOMAIN 5311 5332
FT TRANSMEM 5333 5354
FT DOMAIN 5355 5376
FT TRANSMEM 5377 5398
FT DOMAIN 5399 5420
FT TRANSMEM 5421 5442
FT DOMAIN 5443 5464
FT TRANSMEM 5465 5486
FT DOMAIN 5487 5508
FT TRANSMEM 5509 5530
FT DOMAIN 5531 5552
FT TRANSMEM 5553 5574
FT DOMAIN 5575 5596
FT TRANSMEM 5597 5618
FT DOMAIN 5619 5640
FT TRANSMEM 5641 5662
FT DOMAIN 5663 5684
FT TRANSMEM 5685 5706
FT DOMAIN 5707 5728
FT TRANSMEM 5729 5750
FT DOMAIN 5751 5772
FT TRANSMEM 5773 5794
FT DOMAIN 5795 5816
FT TRANSMEM 5817 5838
FT DOMAIN 5839 5860
FT TRANSMEM 5861 5882
FT DOMAIN 5883 5904
FT TRANSMEM 5905 5926
FT DOMAIN 5927 5948
FT TRANSMEM 5949 5970
FT DOMAIN 5971 5992
FT TRANSMEM 5993 6014
FT DOMAIN 6015 6036
FT TRANSMEM 6037 6058
FT DOMAIN 6059 6080
FT TRANSMEM 6081 6102
FT DOMAIN 6103 6124
FT TRANSMEM 6125 6146
FT DOMAIN 6147 6168
FT TRANSMEM 6169 6190
FT DOMAIN 6191 6212
FT TRANSMEM 6213 6234
FT DOMAIN 6235 6256
FT TRANSMEM 6257 6278
FT DOMAIN 6279 6300
FT TRANSMEM 6301 6322
FT DOMAIN 6323 6344
FT TRANSMEM 6345 6366
FT DOMAIN 6367 6388
FT TRANSMEM 6389 6410
FT DOMAIN 6411 6432
FT TRANSMEM 6433 6454
FT DOMAIN 6455 6476
FT TRANSMEM 6477 6498
FT DOMAIN 6499 6520
FT TRANSMEM 6521 6542
FT DOMAIN 6543 6564
FT TRANSMEM 6565 6586
FT DOMAIN 6587 6608
FT TRANSMEM 6609 6630
FT DOMAIN 6631 6652
FT TRANSMEM 6653 6674
FT DOMAIN 6675 6696
FT TRANSMEM 6697 6718
FT DOMAIN 6719 6740
FT TRANSMEM 6741 6762
FT DOMAIN 6763 6784
FT TRANSMEM 6785 6806
FT DOMAIN 6807 6828
FT TRANSMEM 6829 6850
FT DOMAIN 6851 6872
FT TRANSMEM 6873 6894
FT DOMAIN 6895 6916
FT TRANSMEM 6917 6938
FT DOMAIN 6939 6960
FT TRANSMEM 6961 6982
FT DOMAIN 6983 7004
FT TRANSMEM 7005 7026
FT DOMAIN 7027 7048
FT TRANSMEM 7049 7070
FT DOMAIN 7071 7092
FT TRANSMEM 7093 7114
FT DOMAIN 7115 7136
FT TRANSMEM 7137 7158
FT DOMAIN 7159 7180
FT TRANSMEM 7181 7202
FT DOMAIN 7203 7224
FT TRANSMEM 7225 7246
FT DOMAIN 7247 7268
FT TRANSMEM 7269 7290
FT DOMAIN 7291 7312
FT TRANSMEM 7313 7334
FT DOMAIN 7335 7356
FT TRANSMEM 7357 7378
FT DOMAIN 7379 7400
FT TRANSMEM 7401 7422
FT DOMAIN 7423 7444
FT TRANSMEM 7445 7466
FT DOMAIN 7467 7488
FT TRANSMEM 7489 7510
FT DOMAIN 7511 7532
FT TRANSMEM 7533 7554
FT DOMAIN 7555 7576
FT TRANSMEM 7577 7598
FT DOMAIN 7599 7620
FT TRANSMEM 7621 7642
FT DOMAIN 7643 7664
FT TRANSMEM 7665 7686
FT DOMAIN 7687 7708
FT TRANSMEM 7709 7730
FT DOMAIN 7731 7752
FT TRANSMEM 7753 7774
FT DOMAIN 7775 7796
FT TRANSMEM 7797 7818
FT DOMAIN 7819 7840
FT TRANSMEM 7841 7862
FT DOMAIN 7863 7884
FT TRANSMEM 7885 7906
FT DOMAIN 7907 7928
FT TRANSMEM 7929 7950
FT DOMAIN 7951 7972
FT TRANSMEM 7973 7994
FT DOMAIN 7995 8016
FT TRANSMEM 8017 8038
FT DOMAIN 8039 8060
FT TRANSMEM 8061 8082
FT DOMAIN 8083 8104
FT TRANSMEM 8105 8126
FT DOMAIN 8127 8148
FT TRANSMEM 8149 8170
FT DOMAIN 8171 8192
FT TRANSMEM 8193 8214
FT DOMAIN 8215 8236
FT TRANSMEM 8237 8258
FT DOMAIN 8259 8280
FT TRANSMEM 8281 8302
FT DOMAIN 8303 8324
FT TRANSMEM 8325 8346
FT DOMAIN 8347 8368
FT TRANSMEM 8369 8390
FT DOMAIN 8391 8412
FT TRANSMEM 8413 8434
FT DOMAIN 8435 8456
FT TRANSMEM 8457 8478
FT DOMAIN 8479 8500
FT TRANSMEM 8501 8522
FT DOMAIN 8523 8544
FT TRANSMEM 8545 8566
FT DOMAIN 8567 8588
FT TRANSMEM 8589 8610
FT DOMAIN 8611 8632
FT TRANSMEM 8633 8654
FT DOMAIN 8655 8676
FT TRANSMEM 8677 8698
FT DOMAIN 8699 8720
FT TRANSMEM 8721 8742
FT DOMAIN 8743 8764
FT TRANSMEM 8765 8786
FT DOMAIN 8787 8808
FT TRANSMEM 8809 8830
FT DOMAIN 8831 8852
FT TRANSMEM 8853 8874
FT DOMAIN 8875 8896
FT TRANSMEM 8897 8918
FT DOMAIN 8919 8940
FT TRANSMEM 8941 8962
FT DOMAIN 8963 8984
FT TRANSMEM 8985 9006
FT DOMAIN 9007 9028
FT TRANSMEM 9029 9050
FT DOMAIN 9051 9072
FT TRANSMEM 9073 9094
FT DOMAIN 9095 9116
FT TRANSMEM 9117 9138
FT DOMAIN 9139 9160
FT TRANSMEM 9161 9182
FT DOMAIN 9183 9204
FT TRANSMEM 9205 9226
FT DOMAIN 9227 9248
FT TRANSMEM 9249 9270
FT DOMAIN 9271 9292
FT TRANSMEM 9293 9314
FT DOMAIN 9315 9336
FT TRANSMEM 9337 9358
FT DOMAIN 9359 9380
FT TRANSMEM 9381 9402
FT DOMAIN 9403 9424
FT TRANSMEM 9425 9446
FT DOMAIN 9447 9468
FT TRANSMEM 9469 9490
FT DOMAIN 9491 9512
FT TRANSMEM 9513 9534
FT DOMAIN 9535 9556
FT TRANSMEM 9557 9578
FT DOMAIN 9579 9600
FT TRANSMEM 9601 9622
FT DOMAIN 9623 9644
FT TRANSMEM 9645 9666
FT DOMAIN 9667 9688
FT TRANSMEM 9689 9710
FT DOMAIN 9711 9732
FT TRANSMEM 9733 9754
FT DOMAIN 9755 9776
FT TRANSMEM 9777 9798
FT DOMAIN 9799 9820
FT TRANSMEM 9821 9842
FT DOMAIN 9843 9864
FT TRANSMEM 9865 9886
FT DOMAIN 9887 9908
FT TRANSMEM 9909 9930
FT DOMAIN 9931 9952
FT TRANSMEM 9953 9974
FT DOMAIN 9975 9996
FT TRANSMEM 9997 10000
FT DOMAIN 10001 10022
FT TRANSMEM 10023 10044
FT DOMAIN 10045 10066
FT TRANSMEM 10067 10088
FT DOMAIN 10089 10110
FT TRANSMEM 10111 10132
FT DOMAIN 10133 10154
FT TRANSMEM 10155 10176
FT DOMAIN 10177 10198
FT TRANSMEM 10199 10220
FT DOMAIN 10221 10242
FT TRANSMEM 10243 10264
FT DOMAIN 10265 10286
FT TRANSMEM 10287 10308
FT DOMAIN 10309 10330
FT TRANSMEM 10331 10352
FT DOMAIN 10353 10374
FT TRANSMEM 10375 10396
FT DOMAIN 10397 10418
FT TRANSMEM 10419 10440
FT DOMAIN 10441 10462
FT TRANSMEM 10463 10484
FT DOMAIN 10485 10506
FT TRANSMEM 10507 10528
FT DOMAIN 10529 10550
FT TRANSMEM 10551 10572
FT DOMAIN 10573 10594
FT TRANSMEM 10595 10616
FT DOMAIN 10617 10638
FT TRANSMEM 10639 10660
FT DOMAIN 10661 10682
FT TRANSMEM 10683 10704
FT DOMAIN 10705 10726
FT TRANSMEM 10727 10748
FT DOMAIN 10749 10770
FT TRANSMEM 10771 10792
FT DOMAIN 10793 10814
FT TRANSMEM 10815 10836
FT DOMAIN 10837 10858
FT TRANSMEM 10859 10880
FT DOMAIN 10881 10902
FT TRANSMEM 10903 10924
FT DOMAIN 10925 10946
FT TRANSMEM 10947 10968
FT DOMAIN 10969 10990
FT TRANSMEM 10991 11012
FT DOMAIN 11013 11034
FT TRANSMEM 11035 11056
FT DOMAIN 11057 11078
FT TRANSMEM 11079 11100
FT DOMAIN 11101 11122
FT TRANSMEM 11123 11144
FT DOMAIN 11145 11166
FT TRANSMEM 11167 11188
FT DOMAIN 11189 11210
FT TRANSMEM 11211 11232
FT DOMAIN 11233 11254
FT TRANSMEM 11255 11276
FT DOMAIN 11277 11298
FT TRANSMEM 11299 11320
FT DOMAIN 11321 11342
FT TRANSMEM 11343 11364
FT DOMAIN 11365 11386
FT TRANSMEM 11387 11408
FT DOMAIN 11409 11430
FT TRANSMEM 11431 11452
FT DOMAIN 11453 11474
FT TRANSMEM 11475 11496
FT DOMAIN 11497 11518
FT TRANSMEM 11519 11540
FT DOMAIN 11541 11562
FT TRANSMEM 11563 11584
FT DOMAIN 11585 11606
FT TRANSMEM 11607 11628
FT DOMAIN 11629 11650
FT TRANSMEM 11651 11672
FT DOMAIN 11673 11694
FT TRANSMEM 11695 11716
FT DOMAIN 11717 11738
FT TRANSMEM 11739 11760
FT DOMAIN 11761 11782
FT TRANSMEM 11783 11804
FT DOMAIN 11805 11826
FT TRANSMEM 11827 11848
FT DOMAIN 11849 11870
FT TRANSMEM 11871 11892
FT DOMAIN 11893 11914
FT TRANSMEM 11915 11936
FT DOMAIN 11937 11958
FT TRANSMEM 11959 11980
FT DOMAIN 11981 12002
FT TRANSMEM 12003 12024
FT DOMAIN 12025 12046
FT TRANSMEM 12047 12068
FT DOMAIN 12069 12090
FT TRANSMEM 12091 12112
FT DOMAIN 12113 12134
FT TRANSMEM 12135 12156
FT DOMAIN 12157 12178
FT TRANSMEM 12179 12200
FT DOMAIN 12201 12222
FT TRANSMEM 12223 12244
FT DOMAIN 12245 12266
FT TRANSMEM 12267 12288
FT DOMAIN 12289 12310
FT TRANSMEM 12311 12332
FT DOMAIN 12333 12354
FT TRANSMEM 12355 12376
FT DOMAIN 12377 12398
FT TRANSMEM 12399 12420
FT DOMAIN 12421 12442
FT TRANSMEM 12443 12464
FT DOMAIN 12465 12486
FT TRANSMEM 12487 12508
FT DOMAIN 12509 12530
FT TRANSMEM 12531 12552
FT DOMAIN 12553 12574
FT TRANSMEM 12575 12596
FT DOMAIN 12597 12618
FT TRANSMEM 12619 12640
FT DOMAIN 12641 12662
FT TRANSMEM 12663 12684
FT DOMAIN 12685 12706
FT TRANSMEM 12707 12728
FT DOMAIN 12729 12750
FT TRANSMEM 12751 12772
FT DOMAIN 12773 12794
FT TRANSMEM 12795 12816
FT DOMAIN 12817 12838
FT TRANSMEM 12839 12860
FT DOMAIN 12861 12882
FT TRANSMEM 12883 12904
FT DOMAIN 12905 12926
FT TRANSMEM 12927 12948
FT DOMAIN 12949 12970
FT TRANSMEM 12971 12992
FT DOMAIN 12993 13014
FT TRANSMEM 13015 13036
FT DOMAIN 13037 13058
FT TRANSMEM 13059 13080
FT DOMAIN 13081 13102
FT TRANSMEM 13103 13124
FT DOMAIN 13125 13146
FT TRANSMEM 13147 13168
FT DOMAIN 13169 13190
FT TRANSMEM 13191 13212
FT DOMAIN 13213 13234
FT TRANSMEM 13235 13256
FT DOMAIN 13257 13278
FT TRANSMEM 13279 13300
FT DOMAIN 13301 13322
FT TRANSMEM 13323 13344
FT DOMAIN 13345 13366
FT TRANSMEM 13367 13388
FT DOMAIN 13389 13410
FT TRANSMEM 13411 13432
FT DOMAIN 13433 13454
FT TRANSMEM 13455 13476
FT DOMAIN 13477 13498
FT TRANSMEM 13499 13520
FT DOMAIN 13521 13542
FT TRANSMEM 13543 13564
FT DOMAIN 13565 13586
FT TRANSMEM 13587 13608
FT DOMAIN 13609 13630
FT TRANSMEM 13631 13652
FT DOMAIN 13653 13674
FT TRANSMEM 13675 13696
FT DOMAIN 13697 13718
FT TRANSMEM 13719 13740
FT DOMAIN 13741 13762
FT TRANSMEM 13763 13784
FT DOMAIN 13785 13806
FT TRANSMEM 13807 13828
FT DOMAIN 13829 13850
FT TRANSMEM 13851 13872
FT DOMAIN 13873 13894
FT TRANSMEM 13895 13916
FT DOMAIN 13917 13938
FT TRANSMEM 13939 13960
FT DOMAIN 13961 13982
FT TRANSMEM 13983 14004
FT DOMAIN 14005 14026
FT TRANSMEM 14027 14048
FT DOMAIN 14049 14070
FT TRANSMEM 14071 14092
FT DOMAIN 14093 14114
FT TRANSMEM 14115 14136
FT DOMAIN 14137 14158
FT TRANSMEM 14159 14180
FT DOMAIN 14181 14202
FT TRANSMEM 14203 14224
FT DOMAIN 14225 14246
FT TRANSMEM 14247 14268
FT DOMAIN 14269 14290
FT TRANSMEM 14291 14312
FT DOMAIN 14313 14334
FT TRANSMEM 14335 14356
FT DOMAIN 14357 14378
FT TRANSMEM 14379 14400
FT DOMAIN 14401 14422
FT TRANSMEM 14423 14444
FT DOMAIN 14445 14466
FT TRANSMEM 14467 14488
FT DOMAIN 14489 14510
FT TRANSMEM 14511 14532
FT DOMAIN 14533 14554
FT TRANSMEM 14555 14576
FT DOMAIN 14577 14598
FT TRANSMEM 14599 14620
FT DOMAIN 14621 14642
FT TRANSMEM 14643 14664
FT DOMAIN 14665 14686
FT TRANSMEM 14687 14708
FT DOMAIN 14709 14730
FT TRANSMEM 14731 14752
FT DOMAIN 14753 14774
FT TRANSMEM 14775 14796
FT DOMAIN 14797 14818


```

FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 422 437 BY SIMILARITY.
FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 578 BY SIMILARITY.
FT DISULFID 573 587 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 605 616 BY SIMILARITY.
FT DISULFID 610 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 685 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 723 728 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 760 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 815 824 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 876 891 BY SIMILARITY.
FT DISULFID 893 902 BY SIMILARITY.
FT DISULFID 909 920 BY SIMILARITY.
FT DISULFID 914 929 BY SIMILARITY.
FT DISULFID 931 940 BY SIMILARITY.
FT DISULFID 947 958 BY SIMILARITY.
FT DISULFID 952 967 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.
FT DISULFID 1023 1034 BY SIMILARITY.
FT DISULFID 1028 1043 BY SIMILARITY.
FT DISULFID 1045 1054 BY SIMILARITY.

```

Query Match 10.5%; Score 211.5; DB 1; Length 2437;
 Best Local Similarity 20.7%; Pred. No. 3.5e-08;
 Matches 71; Conservative 21; Mismatches 118; Indels 133; Gaps 13;

```

QY 128 CPGGTGPPDLACGGGSRPCSGNGHCS-----GDSGRQ----- 161
DB 128 CQFGWGGKTCQLADPCGNSPCANGGQSAPESHYICTCPENFHGQTCRODVNECAVSPSP 187
QY 162 -----GDGSCRHMGYOGPLC-----TDCMDG 183
DB 188 CRNGGTCINIVGSYLCRCPPRYTGPHQCRXYQCLPSPCRSGGTQVOTSTTHTCSCLRG 247
QY 184 YFSSLRNETHSICT--ACDESKTCGSLINRDCGCEGVGWVLDGACVVDDECAABPPPC 241

```

```

DB 248 FTGQTCENHVDCTQACENGCGPCIDGINTYNC-HCDKMT-GQYCTEDVDECELSPNAC 305
QY 242 SAAQFCKNANGSYTC-----EECDSS---CVGCTGEGPGNCKE----- 276
DB 306 QNCGTCHNITGGFHCYCVWGWGDDCSENIIDCASACSHGATCDRAVSFCECPHGT 365
QY 277 -----CISGYAREHQOC-----ADVDECSLAECTVCRK 304
DB 366 GILCHLDADACISNPGCKGNSNCDPNVSGKALCTCPRGYGSACNDIDBCSIGANPC-EH 424
QY 305 NENGYTPGSIYVCVCPDGEFEETEDACVPPAEATGES-PTQ 346
DB 425 GGRCLNTKGSFOCKLQGYEG-----PRCEMDVNCCKSNPCQ 461

RESULT 42
FBL5_HUMAN
ID FBL5_HUMAN STANDARD; PRT; 448 AA.
AC Q9UBX5; O75966;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Fibulin-5 precursor (FBLN-5) (developmental arteries and neural crest
  EGF-like protein) (Dance) (Urine p50 protein) (UP50).
GN FBLN5 OR DANCE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Koske G.;
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA MEDLINE-99357779; PubMed-10428823;
RA Nakamura T., Ruiz-Iozano P., Lindner V., Yabe D., Taniwaki M.,
  Furukawa Y., Kobuke K., Tashiro K., Lu Z., Andon N.L., Schaub R.,
  Matsumori A., Saayama S., Chien K.R., Honjo T.;
RT "DANCE, a novel secreted RGD protein expressed in developing,
  atherosclerotic, and balloon-injured arteries.";
RT J. Biol. Chem. 274:22476-22483(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Tissue-urine;
RA Zemel R., Sholto O., Shaul Y.;
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RL [1]
RP FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
  INTERACTION OF INTEGRIN AND THE RGD MOTIF. COULD BE A VASCULAR
  LIGAND FOR INTEGRIN RECEPTORS AND MAY PLAY A ROLE IN VASCULAR
  DEVELOPMENT AND REMODELING.
CC [1] SUBCELLULAR LOCATION: Secreted.
CC [1] TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN HEART, OVARY, AND
  COLON BUT ALSO IN KIDNEY, PANCREAS, TESTIS, LUNG, AND PLACENTA.
  NOT DETECTABLE IN BRAIN, LIVER, THYMUS, PROSTATE, OR PERIPHERAL
  BLOOD LEUCOCYTES.
CC [1] SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation:
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
  CC or send an email to licenselsb@sdb.ch).
CC -----
DB EMBL: AJ133490; CAB38568.1; -
DB EMBL: AF112152; AAD41768.1; -
DB EMBL: AF093118; AAC62107.1; -
DB HSP: P07204; IFGD.
DB MIM: 604580; -

```

DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-like.
 DR Pfam: PF00008; EGF_4.
 DR SMART: SM000179; EGF_CA_4.
 DR SMART: SM00001; EGF_Like_2.
 DR PROSITE: PS00010; ASX_HYDROXYL_4.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 4.
 DR PROSITE: PS01187; EGF_CA_6.
 KW Repeat: EGF-like domain; Calcium-binding; Glycoprotein; Signal.
 FT SIGNAL 1 23
 FT CHAIN 1 23
 FT DOMAIN 24 448
 FT DOMAIN 127 69
 FT DOMAIN 167 167
 FT DOMAIN 168 206
 FT DOMAIN 207 246
 FT DOMAIN 247 287
 FT DOMAIN 288 333
 FT SITE 54 56
 FT DISULFID 131 144
 FT DISULFID 138 153
 FT DISULFID 155 166
 FT DISULFID 172 181
 FT DISULFID 177 190
 FT DISULFID 192 205
 FT DISULFID 211 221
 FT DISULFID 217 230
 FT DISULFID 232 245
 FT DISULFID 251 262
 FT DISULFID 258 271
 FT DISULFID 273 286
 FT DISULFID 292 305
 FT DISULFID 299 314
 FT DISULFID 320 332
 FT CARBOHYD 283 283
 FT CARBOHYD 296 296
 FT CONFLICT 69 70
 FT CONFLICT 147 148
 FT SEQUENCE 448 AA; 50180 MW; 19FCA51FDA328003 CRC64;

Query Match 10.4%; Score 209.5; DB 1; Length 448;
 Best Local Similarity 28.8%; Pred. No. 1e-08;
 Matches 65; Conservative 20; Mismatches 72; Indels 69; Gaps 14;

QY 138 LACGGGSG-----RPGSGNGHSGDG-----SROGDSGCRHMGY---QGPLCTD- 179
 DB 111 LICEFGYQMDSESNOCVVDVDECATDSHOCNPTQICINTEGGYTCSTGQWMLLEGO-CIDI 169
 QY 180 --CNDGYESSLRNETHSICITACDESKCTGSLTRNDCGECEVGVWLDGC--ACVDVDECA 235
 DB 170 DECRYGT-----CQQLCANVPG--SYSC-TCPGFTLNDGSSCDVNECA 212
 QY 236 AEPFGSAOFCCKNANGSYTCSECDSSCGVGTGEGPCKCKISGYAREHS--OCADYDE 293
 DB 213 TENP---CVQTCVNTYSFIC-----RCDPGELEEDGSHGSDMDE 250
 QY 294 CSLAEKTCVKNENGYTPGSYVCVCPDGEFETED--ACVPAARE 337
 DB 251 CSFSEFLC---QHCCVNOPGTYPCSCPGLYLLDNRSCODINECE 293

RESULT 43
 VS41_GIALA
 ID VS41_GIALA STANDARD; PRT; 687 AA.
 AC P92127;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Variant-specific surface protein VSP4A1 precursor (CRISP-90).
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.

OX NCBI_TaxID=5741;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O2-4A1.
 RX MEDLINE=97321554; PubMed=9178264;
 RA Papanastasiou P., Bruderer T., Li Y., Bonmelli C., Koehler P.;
 RT "Primary structure and biochemical properties of a variant-specific
 surface protein of Giardia.";
 RL Mol. Biochem. Parasitol. 86:13-27(1997).
 RP CHARACTERIZATION.
 RX MEDLINE=97233006; PubMed=9078242;
 RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
 RT "The variant-specific surface protein of Giardia, VSP4A1, is a
 RT glycosylated and palmitoylated protein.";
 RL Biochem. J. 322:49-56(1997).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
 CC PLASMA MEMBRANE.
 CC -1- O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
 CC AT THE REDUCING TERMINUS.
 CC -1- PTM: PALMITOYLATED.
 CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: Z83743; CAB06038.1; -
 DR GlycoSuiteDB: P92127; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR SMART: SM00181; EGF_2.
 DR SMART: SM00001; EGF_Like_1.
 DR SMART: SM00261; FU_3.
 KW Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
 KW Signal.
 FT SIGNAL 1 14
 FT CHAIN 15 687
 FT DOMAIN 15 660
 FT TRANSMEM 661 681
 FT DOMAIN 682 687
 FT SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Query Match 10.4%; Score 209.5; DB 1; Length 687;
 Best Local Similarity 23.6%; Pred. No. 1.5e-08;
 Matches 70; Conservative 22; Mismatches 100; Indels 105; Gaps 14;

QY 129 SPGYGPDCILACGGGSGR-PCSGNGHSCGDSRGDSGCRHMGYGP-LCTDCMDGYFS 186
 DB 133 TPSEKSGECILCPDPTDRNGVYANCA-----TCTAPASSTGATCTECMAGTYK 183
 QY 187 SLRNEETHSICITACDESKCTGSLTRNDCGECEVGVWLDGCACVVDVDECAEAPP----- 240
 DB 184 KSDTE---CAACHSDCATCGEANNOCSTGKYLKSNOCVENKNTJNHYPDTSMT 239
 QY 241 -----CSAOF-----CKNAN-----GSYTCSECDSSCGVGTG----- 268
 DB 240 CVACTVLDANCAATCSFDSATKAGKCLCTNSKIRPTITLDGTSTCYE--NSTAGCGAGADNE 297
 QY 269 -----EGPGNCKECISGYAREHGQCAD-- 290
 DB 298 LFMKEDGSACLLCGDPTKEASNDKGVANCRCTCKNANDSPPTCTACLDGYFLERGSCITTC 357
 QY 291 VDEK-SLAETKTCVKNENK-----YNTPGSYVCV-CPDGFEETEDACVPAAREATE 340
 DB 358 ADNCATCSSEATTEEDCKICKAGFLASPGEGKCSISDPTNNGIDGC-----AECTK 409


```

FT DISULFID 729 740 BY SIMILARITY.
FT DISULFID 734 749 BY SIMILARITY.
FT DISULFID 751 760 BY SIMILARITY.
FT DISULFID 767 778 BY SIMILARITY.
FT DISULFID 772 787 BY SIMILARITY.
FT DISULFID 789 799 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 837 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 849 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 908 919 BY SIMILARITY.
FT DISULFID 913 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 946 957 BY SIMILARITY.
FT DISULFID 952 966 BY SIMILARITY.
FT DISULFID 968 977 BY SIMILARITY.
FT DISULFID 984 995 BY SIMILARITY.
FT DISULFID 989 1009 BY SIMILARITY.
FT DISULFID 1011 1020 BY SIMILARITY.
FT DISULFID 1211 1222 BY SIMILARITY.
FT DISULFID 1216 1231 BY SIMILARITY.
FT DISULFID 1233 1242 BY SIMILARITY.
FT DISULFID 1485 1496 BY SIMILARITY.
FT DISULFID 1490 1505 BY SIMILARITY.
FT DISULFID 1507 1516 BY SIMILARITY.
FT DISULFID 1763 1774 BY SIMILARITY.
FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match Best Local Similarity 26.48; Score 209; DB 1; Length 2139;
Matches 63; Conservative 20; Mismatches 78; Indels 78; Gaps 14;

```

OY 125 KVCSPGTYGDFCLACGSGQR-----CSG-NGHSGSGSGSGSGS-CRCHMGYGGP 175
DB 519 KCVCKPSSSGTNCOTSTGDCASALALTPINCNAINGKLNKGTCSMNGTHCYCAVGYSGD 578
OY 176 LC-----TDCMDGYFSSLRNE---THSICGTAC-----DESCCTC 206
DB 579 RCEAENCSPLNCOEP-WVCVQNCCLCPENKYVNCOCATOPONGGECVDPNGDYECCT 637
OY 207 SGLNRCGCEGVEGVWLDGACAPPPCSAOPFKMANGSTCECCDSSCYGC 266
DB 638 RGMWGRTCGN-----DVDECTLHPKICGNG-ICKNEKGSYKC----- 673

```

```

OY 267 TGEPRGNKECISGAREHGOC-ADVDECSLAETCVARNKNCYTPGSYVCVCPDGE 324
DB 674 -----YCTPRTGVH--CDSVDYDC--LSFPC-LNGATCHKKNAYECVCPGE 718

RESULT 45
FBL4_CRIGR STANDARD: PRT: 443 AA.
AC 05058;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE EGF-containing fibulin-like extracellular matrix protein 2 precursor
DE (Fibulin-4) (FBL-4) (H411 protein).
CN EFEMP2 OR FBLN4.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Heine H., Delude R.L., Monks B., Golenbock D.T.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF046870; AAC03101.1; -.
DR HSSP: P00736; IAPQ.
DR InterPro: IPR0000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00008; EGF_4.
DR SMART: SM00179; EGF_CA. 4.
DR SMART: SM00001; EGF_Like; 2.
DR PROSITE: PS00010; ASX_HYDROXYL. 4.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 4.
DR Repeat: PS01187; EGF_CA; 6.
KW Repeat: EGF-like domain; Calcium-binding; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 443
FT -----
FT DOMAIN 36 81
FT DOMAIN 123 163
FT DOMAIN 164 202
FT DOMAIN 203 242
FT DOMAIN 243 282
FT DOMAIN 283 328
FT DOMAIN 329 368
FT DISULFID 127 140
FT DISULFID 134 149
FT DISULFID 151 162
FT DISULFID 168 177
FT DISULFID 173 186
FT DISULFID 188 201
FT DISULFID 207 217
FT DISULFID 213 226
FT DISULFID 228 241
FT DISULFID 247 258
FT DISULFID 254 267
FT DISULFID 269 281
FT DISULFID 287 300
FT DISULFID 294 309
FT DISULFID 315 327

```

FT CAROHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 443 AA; 49432 MW; 0BCFE5D7323D9E5F CRC64;

Query Match 10.4%; Score 208; DB 1; Length 443;
 Best Local Similarity 37.0%; Pred. No. 1.3e-08;
 Matches 47; Conservative 12; Mismatches 56; Indels 12; Gaps 5;

OY 216 ECEVGG--WVLDGACVYDVECAAEPPPCSAQPCNKANSYTCEDSCVG-CTEGGP- 271
 DB 38 ECTDGTGEMDADSHCRDYNDECLTTPACKGEMKCNHGYGLCLPSAAVINLHGEGR 97
 OY 272 -----GNCKECISGY-AREHGQCADVDECSIAEKTQVRKNECNYTPGSSYVCVCPDGEF 324
 DB 98 PVPRAQHNPCCPPGYPEPDEQESCVYDVECAQALHDC-RPSQCHNLPGSYQCTCPDGYR 156
 OY 325 ETEDACV 331
 DB 157 KVGPECV 163

RESULT 46
 FBIL_MOUSE
 ID FBIL_MOUSE STANDARD; PRT; 443 AA.
 AC Q9WVJ9;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE EGF-containing fibulin-like extracellular matrix protein 2 precursor
 DE (Fibulin-4) (FBLN-4) (Mutant p53 binding protein 1).
 GN EFEMP2 OR FBLN OR MBP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=9308589; PubMed=10380882;
 RA Gallagher W.M., Argentini M., Sierra V., Bracco L., Debussche L.,
 RA Conseiller E.;
 RT "MBP1: a novel mutant p53-specific protein partner with oncogenic
 RT properties.";
 RL Oncogene 18:3608-3616(1999).
 CC -I- SUBUNIT: BINDS PREFERENTIALLY TO P53 MUTANTS.
 CC -I- SUBCELLULAR LOCATION: Secreted.
 CC -I- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF104223; AAD45219.1; .
 DR HSSP; P00736; IAPQ.
 DR MGD; MGI:1891209; Efemp2.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF-Ca.
 DR InterPro; JPR001491; Thrombomodulin.
 DR Pfam; PF00008; EGF; 4.
 DR PRINTS; PR00907; THROMMODULN.
 DR SMART; SM00179; EGF_CA; 4.
 DR SMART; SM00001; EGF_Like; 2.
 DR PROSITE; PS00022; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS01187; EGF_CA; 6
 KW Repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.

FT SIGNAL 1 25
 FT CHAIN 26 443
 FT FT
 FT DOMAIN 36 81
 FT DOMAIN 123 163
 FT DOMAIN 164 202
 FT DOMAIN 203 242
 FT DOMAIN 243 282
 FT DOMAIN 283 328
 FT DISULFID 127 140
 FT DISULFID 134 149
 FT DISULFID 151 162
 FT DISULFID 168 177
 FT DISULFID 173 186
 FT DISULFID 188 201
 FT DISULFID 207 217
 FT DISULFID 213 226
 FT DISULFID 228 241
 FT DISULFID 247 258
 FT DISULFID 254 267
 FT DISULFID 269 281
 FT DISULFID 287 300
 FT DISULFID 294 309
 FT DISULFID 315 327
 FT CAROHYD 198 198
 FT CAROHYD 394 394
 SQ SEQUENCE 443 AA; 49425 MW; 4969C0328A23DD88 CRC64;

Query Match 10.3%; Score 207; DB 1; Length 443;
 Best Local Similarity 37.0%; Pred. No. 1.6e-08;
 Matches 47; Conservative 11; Mismatches 57; Indels 12; Gaps 5;

OY 216 ECEVGG--WVLDGACVYDVECAAEPPPCSAQPCNKANSYTCEDSCVG-CTEGGP- 271
 DB 38 ECTDGTGEMDADSHCRDYNDECLTTPACKGEMKCNHGYGLCLPSAAVINLHGEGR 97
 OY 272 -----GNCKECISGY-AREHGQCADVDECSIAEKTQVRKNECNYTPGSSYVCVCPDGEF 324
 DB 98 PPAHAHQNPCCPPGYPEPDEQESCVYDVECTQALHDC-RPSQCHNLPGSYQCTCPDGYR 156
 OY 325 ETEDACV 331
 DB 157 KVGPECV 163

RESULT 47
 FP2_MYTGA
 ID FP2_MYTGA STANDARD; PRT; 473 AA.
 AC Q25464;
 DT 30-MAY-2000 (Rel. 39; Created)
 DT 30-MAY-2000 (Rel. 39; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Adhesive plaque matrix protein 2 precursor (Foot protein 2) (MGFP2)
 DE (MGFP-2).
 GN FP2.
 OS Mytilus galloprovincialis (Mediterranean mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
 OC Mytiloidea; Mytilidae; Mytilus.
 OX NCBI_TaxID=29158;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Foot;
 RC MEDLINE=95204464; PubMed=7896812;
 RX Inoue K., Takeuchi Y., Miki D., Odo S.;
 RA "Mussel adhesive plaque protein gene is a novel member of epidermal
 RT growth factor-like gene family.";
 RL J. Biol. Chem. 270:6698-6701(1995).
 CC -I- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC -I- SUBCELLULAR LOCATION: Secreted.

Wed Sep 11 09:09:34 2002

us-09-905-075-2.top50.rsp

Page 61

Search completed: September 10, 2002, 11:13:02
Job time: 219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 11:08:33 ; Search time 20.91 Seconds
(without alignments)
1622.168 Million cell updates/sec

Title: US-09-905-075-2
Perfect score: 2005
Sequence: 1 MRLPRRALGLPLLLLP.....AEAEATGEGSPQLSPREDL 353

Scoring table: BLOSUM62, Gapop 10.0, Capext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 65 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	959.5	47.9	417	2 T08724	hypothetical prote
2	540	26.9	356	2 T20656	hypothetical prote
3	270	13.5	2871	2 A55567	fibritillin I - bov
4	268.5	13.4	3002	2 A47221	fibritillin I precu
5	268	13.4	1820	2 A55494	latent transformin
6	263	13.1	1548	2 S34563	serine proteinase
7	254.5	12.7	601	2 B36346	fibulin 1 precurs
8	254.5	12.7	683	2 C36346	fibulin 1 precurs
9	253.5	12.6	685	2 S78040	fibulin, splice fo
10	253.5	12.6	705	2 S34968	hypothetical prote
11	252.5	12.6	3507	2 T34513	hypothetical prote
12	252.5	12.6	2871	2 A55624	fibritillin-1 precu
13	252.5	12.6	2907	2 A57278	fibritillin-2 precu
14	250	12.5	3623	2 T09456	intrinsic factor-B
15	249.5	12.4	2918	2 A54105	fibritillin-2 precu
16	248	12.4	915	1 A48225	subtilisin-like pr
17	247	12.3	675	1 KXRT5	plasma protein S p
18	247	12.3	1221	2 A49457	latent transformin
19	246.5	12.3	1251	2 A57293	hypothetical prote
20	245.5	12.2	3871	2 T22812	hypothetical prote
21	242.5	12.1	1184	2 A55184	transforming growt
22	240.5	12.0	1394	2 A35626	plasma protein S p
23	239	11.9	676	1 KXRT5	probable proteol
24	239	11.9	915	2 B48225	subtilisin-like pr
25	239	11.9	1299	2 T43251	subtilisin-like pr
26	236	11.8	899	2 G02428	subtilisin-like pr
27	236	11.8	915	2 JC6148	hypothetical prote
28	234	11.7	1106	2 T18739	plasma protein S p
29	232	11.6	675	1 KXRT5	

ALIGNMENTS

30	232	11.6	1376	2 G00043	osteonidogen - hu
31	231.5	11.5	1069	2 T42681	hypothetical prote
32	228.5	11.4	642	2 S3434	plasma protein S p
33	228	11.4	1680	2 A3434	furin (EC 3.4.21.7
34	227	11.3	2524	2 A35844	notch protein - Af
35	226	11.3	2531	2 A46019	Notch-1 protein -
36	225.5	11.2	646	2 S38819	plasma protein S -
37	224.5	11.2	1574	2 T13954	MEGF6 protein - ra
38	224	11.2	1220	2 A56136	tagged protein pre
39	222.5	11.1	886	2 A57172	probable hormone r
40	222	11.1	675	1 KXRT5	plasma protein S p
41	222	11.1	838	2 T20125	hypothetical prote
42	222	11.1	1712	2 A38261	masking protein pr
43	221	11.0	1203	2 A49175	Notch B protein -
44	221	11.0	1247	1 MMHND	Notch B protein -
45	221	11.0	2555	2 A40043	Notch protein precu
46	220	11.0	2531	2 S1818	Notch protein homo
47	220	11.0	3623	2 T08618	intrinsic factor-B
48	218	10.9	2471	2 A49128	cell-fate determin
49	217.5	10.8	1357	2 T16860	hypothetical prote
50	217.5	10.8	3672	2 T23433	hypothetical prote
51	217.5	10.8	3704	2 T37316	probable lamnin a
52	216.5	10.8	2703	1 A24420	notch protein - fr
53	214	10.7	2321	2 S78549	notch3 protein - h
54	213.5	10.6	1245	1 MMMSND	Notch3 protein - h
55	213	10.6	642	2 S3433	plasma protein S p
56	213	10.6	673	2 A48089	growth arrest-spec
57	212	10.6	2318	2 S45306	growth 3 protein -
58	211.5	10.5	2437	2 S42612	transmembrane prot
59	210.5	10.5	1620	2 T27283	hypothetical prote
60	210	10.5	674	2 T55476	growth potentiatio
61	209	10.4	1766	2 A42125	trophozoite cystel
62	209	10.4	2139	2 A35672	crumbs protein - f
63	205.5	10.2	473	2 A56175	adhesive plaque pr
64	204.5	10.2	810	2 T10756	Nel-homolog protei
65	204.5	10.2	1964	2 T09059	notch4 - mouse

RESULT 1	1	hypothetical protein DKFp566d213.1 - human
T08724	1	
C:Species: Homo sapiens (man)		
C>Date: 11-Jun-1999 #sequence-revision 11-Jun-1999 #text-change 13-Aug-1999		
C:Accession: T08724		
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.		
submitted to the Protein Sequence Database, May 1999		
A:Accession: T08724		
A:Molecule type: mRNA		
A:Residues: 1-417 <KOE>		
A:Cross-references: EMBL:AL050275		
A:Experimental source: fetal kidney; clone DKFp566d213		
C:Genetics:		
A>Note: DKFp566d213.1		
Query Match	47.9%	Score 959.5; DB 2; Length 417;
Best Local Similarity	47.9%	Pred. No. 3.1e-54;
Matches 169; Conservative 47; Mismatches 118; Indels 19; Gaps		
OY 7 AALGLPLLLLP-----PAPBA--KKPTPCRCRGLVDFKNGMVDPAKKNFGG 56		
DB 12 AVLWGLSLFLNDPGPIWLPSPSPPOPPHPCHTCRGLVDSFNKGLERTIRDNFGG 71		
OY 57 NTAMEEKLTSYESESEIRLTLEIGCSSDPECNOMLEAQEHLEAMWLQKSEYDPLF 116		
DB 72 NTAMEEMLSKTKDSEIRLVEVLSGVCSKSPBCHRLLESEELVESWVFRKQCAPLDF 131		
OY 117 EMFCVKTLKVCSPGYGPDCLACGGGQRCPCSGNGHCSGDSRGDSCRCMHGYGQPL 176		

Db 132 QWLCSDSLKCCPAGTFGSPCLPFGGTERPCGGYQCGEGTRGSGHCDCAQAGYGEA 191
 QY 177 CTFCDMGFSSSLNETHSICITACDESKCTCSGLTNRDCGCEYGVWLDDEGACVDVDECA 236
 Db 192 CGCGGAGTGEAEKNASHLYCSACFGPCARCSGSEESNCLDCKGMALHLKCTVDIECGT 251
 QY 237 EPPPCSAOFCFKNANSYTCCECDSSCVGCTGEGPNCCKEJCISGVAREHGOACADVDECSL 296
 Db 252 EGANCGADQFCVNTGSEYECRDCAKACLGCMGAPGRCKKCSFGYQVGSKCLDVDECE- 310
 QY 297 AETCYCRKNKNCNTPGSYVCVCPDGFEEEDPCVP---PAEA---EATEGE 342
 Db 311 -TEVCGENKQCCENTEGYRCICAECKQMEGICVKEQIPESAGFSEMTED 362

RESULT 2

hypothetical protein F09E8.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20656
 R:Percy, C.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z19307
 A:Accession: T20656
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-356 <TIL>
 A:Cross-references: EMBL:Z73896; PIDN:CAA98055.1; GSPDB:GN00022; CESP:F09E8.2
 A:Experimental source: clone F09E8
 C:Genetics:
 A:Gene: CESP:F09E8.2
 A:Map position: 4
 A:Introns: 39/3; 67/2; 99/3; 134/1; 165/3; 208/1; 241/1; 287/1

Query Match 26.9%; Score 540; DB 2; Length 356;
 Best local Similarity 36.1%; Pred. No. 1.5e-27;
 Matches 108; Conservative 43; Mismatches 110; Indels 38; Gaps 6;

QY 11 LPFLLLLPAPPAKKPRPCHRCRGIVDKFNGMGVDTAKKNGGNTAMEEKLKYES 70
 Db 7 LIAVLIGANSQKEVTITKNEKRCFNPVSTFDGLKTAHNFAGDITAMEEKNLKYKT 66
 QY 71 SEIRLEILEJGCESS-----DFECNMLEAEQEBHLBAMWLKSEYDLP 116
 Db 67 SETRIIEVLEGVCKKSLPMDNFMGLAIEFEKCSQLEKHEETIEFY--YNOGINNMS 124
 QY 117 EMFCVTKLVKCSPTGYGPDCLACGGGSGRP--CSGNGHCSGDSRQGDSCRCRHMGYOG 174
 Db 125 NMVLCEQQLKICPDGHFGKNCCEQCPGISERKADVCFKSGCHGDSREGSGKCKCTGYTG 184
 QY 175 PLCTDMDGFFSSLRNETHSICITACDESC-KTCSGLTNRCCGCEYGVWLDDEGACVDVE 233
 Db 185 NLCRICDIEFFESKRVQGVCKKCHGCGICVSSSESSKCSCKKMGWKLTEGSCADVNE 244
 QY 234 CAAPPPCSAOPCKNANSYTCCECDSSCVGCTGEGPNCCKEJCISGVAREHGOACADVD 292
 Db 245 CQNESACTKHEICVMTVGSFKCE-----CKE---GYKKDDQDNQOPD 284

RESULT 3

fibrillin 1 - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000
 C:Accession: A55567
 R:Ristiera, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
 Genomics 23, 480-485, 1994
 A:Title: Sequence of the coding region of the bovine fibrillin cDNA and localization to
 A:Reference number: A55567; MUID:95137557
 A:Accession: A55567
 A:Status: preliminary

A:Molecule type: mRNA
 A:Residues: 1-2871 <TIL>
 A:Cross-references: GB:I28748; NID:9508427; PIDN:AAA74122.1; PID:9508428
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:1201-1236/Domain: EGF homology <EGF>

Query Match 13.5%; Score 270; DB 2; Length 2871;
 Best local Similarity 24.8%; Pred. No. 1.3e-09;
 Matches 107; Conservative 33; Mismatches 128; Indels 164; Gaps 25;

QY 19 PPAPPAKKPRPC-----HRCGLYDKF---NQGAVDTAKNFGGNTA 59
 Db 1141 PPGHOLAPNISACIDINECELSAHLCPHRCVNLTKYOCACNPYHSPTDILF----- 1194
 QY 60 WKEKTLSEKSSERIRLEILEGCS-----SDEFCN-----QMLEAGEHLEAMWLQ 108
 Db 1195 -----CVIDECSLTMNGCETFTNSREGYSCQPGFALMPDOR----- 1234
 QY 109 KSEYDLPFEMFCVTKLVKVC---CS--PGTYGPDCLACOG---GSORPCSGNGHC--- 154
 Db 1235 --SCTDIDE--CEDNPNICDGGGCTNIPGEY--RCICLYGFMASEDMKCTVDVNECDLNP 1288
 QY 155 ---SGD--GSROGDGSCRCRHMGYOGLP---CTDCMDGYFSSLRNETHSICITACDESKCT 205
 Db 1289 NCLSGTCENTGKSFYCHCDMGYSKKKRTGCTDINECEIGHNCDRAHVAVTNTAGSFK- 1347
 QY 206 CSGLTNRDCGCEYGVWLDDEGACVDVDECAAPPPCSAOPCKNANSYTC----- 256
 Db 1348 CS-----CSPGWIGIGIKCTDIDECNSGTHMCSQAHADCKNTMGSTRICLCKEYTG 1398
 QY 257 -----EBC-----DSCVCGCTG-----EGP----- 271
 Db 1399 GFTCTDLDCSENLNLGNGQCLINAPGYRCEDCMGFVSPADKACEDIDECSLPNICVF 1458
 QY 272 GNCK-----ECISGYA--REHGOACADVDECSLAETCYCRKNKNCYTPGSYVCVD 321
 Db 1459 GTCHNDPLGFEKCECEYELDRSGGCTDVNRC-IDPTTCI--SGNCVMTPGSYTCDCPP 1515
 QY 322 GFE--ETEDACV 331
 Db 1516 DPELANPTRVGCV 1527

RESULT 4

fibrillin 1 precursor - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 02-Jun-1995 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
 C:Accession: A47221; I54355; S17064; I59574; S17062; S62111; A34198
 R:Corson, G.M.; Chabberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y.
 Genomics 17, 476-484, 1993
 A:Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain stru
 A:Reference number: A47221; MUID:94010947
 A:Accession: A47221
 A:Molecule type: mRNA
 A:Residues: 1-337, 'T', 339-1029 <COR>
 A:Cross-references: GB:X63556
 R:Perleir, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Panglman, T.; Bo
 Hum. Mol. Genet. 2, 961-968, 1993
 A:Title: Genomic organization of the sequence coding for fibrillin, the defective gen
 A:Reference number: I54355; MUID:93372860
 A:Accession: I54355
 A:Molecule type: mRNA
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Residues: 132-3002 <PER>
 A:Cross-references: GB:LI3923; NID:9306745; PIDN:AA02036.1; PID:9306746
 R:Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y.
 Nature 352, 334-337, 1991
 A:Title: Partial sequence of a candidate gene for the Marfan syndrome.
 A:Reference number: S17064; MUID:91304568
 A:Accession: S17064
 A:Molecule type: mRNA

A:Residues: 1030-3002 <MAS>
A:Cross-references: EMBL:X63556
R:Dieltz, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J.
Science 259, 680-683, 1993
A:Title: The skipping of constitutive exons in vivo induced by nonsense mutations.
A:Reference number: 159574; PMID:93157831
A:Accession: 159574
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 2217-2288, 'I', 2290-2325 <RES>
A:Cross-references: GB:S54426; NID:g264860; PIDN:AAB25244.1; PID:g264861
R:Lee, B.; Godfrey, M.; Vitale, E.; Horl, H.; Mattel, M.G.; Sarfarazi, M.; Tsipouras, P.
Nature 352, 330-334, 1991
A:Title: Linkage of Marfan syndrome and a phenocypically related disorder to two different
A:Reference number: 517062; MUID:91304567
A:Accession: 517062
A:Molecule type: mRNA
A:Residues: VLVTVTFLEYSNKMV, 944-1444 <LEE1>
A:Cross-references: EMBL:X62008; NID:931398; PIDN:CAB56534.1; PID:g5924015
A:Accession: 562111
A:Molecule type: protein
A:Residues: 1166-1176, 'X', 1178-1180, 'D', 1182-1185 <LEE2>
R:Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.
J. Biol. Chem. 264, 21381-21385, 1989
A:Title: Connective tissue microfibrils. Isolation and characterization of three large F
A:Reference number: A34198; MUID:90078246
A:Accession: A34198
A:Molecule type: protein
A:Residues: 565-575, 1890-1892, 'I', 1894-1900 <MAD>
C:Comment: Fibrillin is a major component of elastin-associated microfibrils.
C:Genetics:
A:Gene: GDB:FBNI
A:Cross-references: GDB:127115; OMIM:134797; OMIM:154700
A:Map position: 15q21.1-15q21.1
A:Introns: 2236/1; 2258/1; 2297/1
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein; M
F.1-3002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicted
F.1332-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MAD>
F.1332-1367/Domain: EGF homology <EGF>
F.1457-1492/Domain: EGF homology <EGF2>
F.2262-2295/Domain: EGF homology <EGF1>

[illegible]

Db 1616 CTDVNEC-LDPPTCI--SGNCVNTPGSTICDCPPPELNPTRVGCV 1658

RESULT 5

A55494

latent transforming growth factor-beta-binding protein - human

C:Species: Homo sapiens (man)

C:Date: 03-Mar-1995 #sequence-revision 03-Mar-1995 #text-change 11-Jan-2000

C:Accession: A55494

J:Moren, A.; Olafsson, A.; Stenman, G.; Sahlin, P.; Kanzaki, T.; Claesson-Welsh, L.; J. Biol. Chem. 269, 32469-32478, 1994

A:Title: Identification and characterization of LTBP-2, a novel latent transforming growth factor-beta-binding protein

A:Reference number: A55494; MUID:95096101

A:Accession: A55494

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1820 <MR>

A:Cross-references: GB:237976

C:Genetics:

A:Gene: GDB:LTBP2

A:Cross-references: GDB:568901

A:Map position: 11pter-11qter

C:Superfamily: unassigned EGF-related proteins; EGF homology

F:1222-1257/Domain: EGF homology <EGF>

F:1525-1565/Domain: EGF homology <EGF1>

Query Match 13.4%; Score 268; DB 2; Length 1820;

Best Local Similarity 25.4%; Pred. No. 1,2e-09;

Matches 90; Conservative 24; Mismatches 94; Indels 146; Gaps 19;

```

Query Match          13.4%; Score 268; DB 2; Length 1820,
Best Local Similarity 25.4%; Pred. No. 1.2e-09;
Matches    90; Conservative   24; Mismatches   94; Indels   146; Gaps   19;

Oy      113 PDLEFMFCVKTLLKVCSPG-----TYGPDCIACOGSGQRPCSGNGHCSGDGSROG 162
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       865 PDGYR-CV-----CSPGYQLHPQAYCTDDNECL-----RDPCQGKGRCI---NRVG 907
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      163 DSCGRCHNGY-----GPRLCI-T-----DGMGFPSRLANE 191
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       908 SYSCFCEPGYTLAVNSGAATQECODINECPQGVCSGGCTNTGEGSYHCSDDGITLVNRRKH 967
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      192 THSI-----CTACDESKCTCGS-----LTNRDCG----- 215
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       968 CODINECHRPHGCPDPGRCVNSPFGSYTCLACEGVRGGSGCDVDNECLLTPGCALHGKTN 1027
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      216 -----ECEVNGWVL--DEGACVDVDECAAEPPPSAAOFCKNANGSYTCBEOCSS---- 262
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1028 LEGSFRCSCGEGYEVTDSBEKGGODVDECASF-RASCPTG-LCIINTGESSFACSCAENGWVN 1085
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      263 -----CVG--CTG--EGRPANKEICISGAR--HGOCADVDECIAEKT 300
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1086 EDGTACEDLDDECAFPGVCPSPGVCITNTAGSFSCKDGDGTRSPPLDSEDVEDDECDDPSS 1145
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      301 CVRKNNENCYTMPGSYYVCVCPDGFEEET-----EDACVPRAEAETEG 341
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1146 CL--GGECKKNMGYSYOCLCPQGFOLANGTVCEDVNECMGEHEHCAHGLCLNSHG 1197
           || : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT        6
S34583
serine proteinase (EC 3.4.21.-) PCGB - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C:Accession: S34583
R:Nakagawa, T.; Murakami, K.; Nakayama, K.
FEBS Lett. 327: 165-171, 1993
A>Title: Identification of an isoform with an extremely large Cys-rich region of PC6,
A:Reference number: S34583; MUID:93327934
A:Accession: S34583
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1548 <NA>
A:Cross-references: GB:D17583; NID:g407344; PIDN:BAA04507.1; PID:d1005033; PID:g44037
C:Keywords: hydrolase; serine proteinase

```


Db 291 KLOCKSGEFDALGNCIDINECLISAPCPIGHNCTINTGSEYTCOK----- 336
Oy 273 NCKECISGY-AREHG-OCADVDECSLAECTVKRNENYNTPGSYVCPDG--FEETED 328
Db 337 NVPCGKGYHLNEGTCTRCVDDECAAPAEPC-GKHRCVNSPGRCECKTGYTFDGTGR 395
Oy 329 ACVPAEAEATEG 341
Db 396 MCVNNECORRYPG 408
RESULT 9
S78040
fibulin, splice form C precursor - mouse
N:Alternate names: basement-membrane protein BM-90
C:Species: Mus musculus (house mouse)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 23-Mar-2001
C:Accession: S78040; S78560; S36440
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
Eur. J. Biochem. 215, 733-740, 1993
A:Title: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dependent
A:Reference number: S34968; MUID:93358897
A:Accession: S78040
A:Molecule type: mRNA
A:Residues: 1-685 <PAN>
A:Cross-references: EMBL:X70854
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
submitted to the EMBL Data Library, January 1993
A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dep
A:Reference number: S36440
A:Accession: S78560
A:Molecule type: mRNA
A:Residues: 1-39, 'P', '41-685 <CHU>
A:Cross-references: EMBL:X70854
C:Genetics:
A:Introns: 568/3
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: alternative splicing; basement membrane; calcium binding; extracellular mat
F:1-29/Domin: signal sequence #status predicted <SIG>
F:30-685/Product: fibulin, splice form C #status predicted <MAT>
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 12.6%; Score 253.5; DB 2; Length 685;
Best Local Similarity 27.1%; Pred. No. 4.6e-09;
Matches 80; Conservative 24; Mismatches 108; Indels 83; Gaps 14;
Oy 115 LFEMFCVTKLVCCSPGTGGPCLACOGS--QRP-----CSGNHCSG 156
Db 131 MISYCGGLVFRACCVKARENSDFVGNAGADLPDPAKIPDEEDQDPYLNDRCGGGPKQ 190
Oy 157 DGSROGDC-SCRCMHGYO---GPICTD---CMGYSF---SLRNETHSICTACDSC 203
Db 191 QCRDTGDEVITSCFVGYLOSGVSCEDINECTGSHNCRLESCINTVGSFRCORDSC 250
Oy 204 KTCGSLTN---RDGCEEV-----GWVLDG-GACVD 230
Db 251 GTGYELTDDNNCKDIDECETGHNCPDFICONTLGSFRCPKLOCKSGFIODALGNCID 310
Oy 231 VDECAEPPPCSAOFCFNANGSYTCCECDSSCVGCTGEGPNCKECISGY-AREHG-QC 288
Db 311 INECISISAPCPVGTCTINTGSEYTCOK-----NVNCGRGYHLNEGTGR 356
Oy 289 ADVDECSLAECTVKRNENYNTPGSYVCPDG--FEETEDACVPAEATEG 341
Db 357 VDVECAPAEPC-GKHHCNLSPSGFRCECKAGFYFDGISRTCYDINECORRYPG 410
RESULT 10
S34968
fibulin, splice form D precursor - mouse
N:Alternate names: basement-membrane protein BM-90; calcium-binding protein BM-90
C:Species: Mus musculus (house mouse)

C:Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 31-Jan-2000
C:Accession: S34968; S36441; S13814
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
Eur. J. Biochem. 215, 733-740, 1993
A:Title: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-depend
A:Reference number: S34968; MUID:93358897
A:Accession: S34968
A:Molecule type: mRNA
A:Residues: 1-705 <PAN>
R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
submitted to the EMBL Data Library, January 1993
A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-
A:Reference number: S36440
A:Accession: S36441
A:Molecule type: mRNA
A:Residues: 1-39, 'P', '41-705 <PAN>
A:Cross-references: EMBL:X70854; NID:9396820; PIDN:CA50207.1; PID:9396821
A:Experimental source: Cell-line F9 teratocarcinoma
R:Kluge, M.; Mann, K.; Dziadek, M.; Timpl, R.
Eur. J. Biochem. 193, 651-659, 1990
A:Title: Characterization of a novel calcium-binding 90-kDa glycoprotein (BM-90) shar
A:Reference number: S13814; MUID:91065369
A:Accession: S13814
A:Molecule type: protein
A:Residues: 28:31-49, 'X', '51-53, 'XX', '110-117, '231-240, 'X', '242-243, '339-362, 'S', '364-387, '4
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: alternative splicing; basement membrane; calcium binding; extracellular m
F:1-29/Domin: signal sequence #status predicted <SIG>
F:30-705/Product: fibulin, splice form D #status predicted <MAT>
F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 12.6%; Score 253.5; DB 2; Length 705;
Best Local Similarity 27.1%; Pred. No. 4.7e-09;
Matches 80; Conservative 24; Mismatches 108; Indels 83; Gaps 14;
Oy 115 LFEMFCVTKLVCCSPGTGGPCLACOGS--QRP-----CSGNHCSG 156
Db 131 MISYCGGLVFRACCVKARENSDFVGNAGADLPDPAKIPDEEDQDPYLNDRCGGGPKQ 190
Oy 157 DGSROGDC-SCRCMHGYO---GPICTD---CMGYSF---SLRNETHSICTACDSC 203
Db 191 QCRDTGDEVITSCFVGYLOSGVSCEDINECTGSHNCRLESCINTVGSFRCORDSC 250
Oy 204 KTCGSLTN---RDGCEEV-----GWVLDG-GACVD 230
Db 251 GTGYELTDDNNCKDIDECETGHNCPDFICONTLGSFRCPKLOCKSGFIODALGNCID 310
Oy 231 VDECAEPPPCSAOFCFNANGSYTCCECDSSCVGCTGEGPNCKECISGY-AREHG-QC 288
Db 311 INECISISAPCPVGTCTINTGSEYTCOK-----NVNCGRGYHLNEGTGR 356
Oy 289 ADVDECSLAECTVKRNENYNTPGSYVCPDG--FEETEDACVPAEATEG 341
Db 357 VDVECAPAEPC-GKHHCNLSPSGFRCECKAGFYFDGISRTCYDINECORRYPG 410
RESULT 11
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favell, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:
 A:Gene: CESP:2K783.1
 A:Map position: 3
 A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
 3504/1

Query Match 12.6%; Score 253; DB 2; Length 3507;
 Best Local Similarity 19.7%; Pred. No. 1.8e-08;
 Matches 103; Conservative 41; Mismatches 100; Indels 280; Gaps 23;

QY 83 CESSDEPCOMLEAOEHLLEAMWLKSEYPDLEFMEVCYKTK----- 125
 DB 1454 CATGCHNEHESARCO-----NYVGYACFCPTGFRKADGSCODIDECTEHN 1500
 QY 126 -VCC-----SPGTGPDG-----LACGGGSGRPC-----SGNGHSG----- 156
 DB 1501 STCCGANAKCVKPGTYSCCEENGFLGDGYQCVPTTKPCDSTQSKSHCSSESNSCEVD 1560
 QY 157 --DGS-----RQDGSRCRHMGYGP- 175
 DB 1561 TWDGVECKEKGKSGKVCEDINECYAEKAPCSLNNVNMGTFCSCCKQGRGDG 1620
 QY 176 -LCTD-----C-MGYFSSLNETHS-----ICT-----ACDESK 204
 DB 1621 FMCITDINECDERHPCHPNAECTNLEB---SFKCECHSGFEGDGKIKCTNPLERSCEDEVER 1677
 QY 205 TCSGLNRPD-----GCEVGMVLDE--GACVDYDECAAEPPPCS-AAQPFCK 248
 DB 1678 FCGRDVHSCSLVRIYNGSLSVSCCEPGRFREKESNSCVDIDECESRNNDPASAVCV 1737
 QY 249 NANGSYTCB-----ECDSGVC-----TGECP 271
 DB 1738 NTEGSRCECARGYEGEGVCTDIDECRGMAGCDMAMCINRMGSCGCKMAGYTGDA 1797
 QY 272 -----GNCK----- 275
 DB 1798 TCIKIEEPKSDKTACTDEMSRLCELEKQCTVDEBEVPOGACLPGHHPINTCQSLQI 1857
 QY 276 -----ECI-----SGYAREHGOCADYDECSLAKTCVARKENCY 309
 DB 1858 SGLCAQKNCNKNAECIDIHPSHSCSPDGFIDGMITDDVDECCNNA-GMCDDEMTKE 1916
 QY 310 NTPGSYVCVPDGFEEETEDACVPAPAEATEGESPTQLPSREDL 353
 DB 1917 NTIGSPNCVCLLEGFKKVDKCVYDEKKQ-----ENRERI 1950

RESULT 12

A55624
 fibrillin-1 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 11-Jan-2000
 C:Accession: A55624
 R:Ylin, W.; Smiley, E.; Germiller, J.; Sanguinetti, C.; Lawton, T.; Pereira, L.; Ramirez, J. Biol. Chem. 270, 1798-1806, 1995
 A:Title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin 96
 A:Reference number: A55624; MUID:95130561
 A:Accession: A55624
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2871 <YIN>
 A:Cross-references: GB:L29454; NID:9575509; PIDN:AAA56840.1; PID:9575510
 C:Genetics:
 A:Gene: Fbn-1
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:1201-1236/Domain: EGF Homology <EGF>

Query Match 12.6%; Score 252.5; DB 2; Length 2871;
 Best Local Similarity 25.1%; Pred. No. 1.6e-08;
 Matches 102; Conservative 31; Mismatches 124; Indels 149; Gaps 24;

QY 30 PCHRCGLVDKF-----NQMVDYTAKNFGGNGTAMEEKTLSKYESSEIRLLLEGLCES 85
 DB 1167 PHGRCVNLIGKYQACANPGYHPTHDLF-----CVDIECSIMNGCET 1210
 QY 86 -----SDECN-----QMLEAQEHLLEAMWLKSEYPDLEFMEVCYKTKC-----CS- 129
 DB 1211 FCTNSDGSYECSCQPGFALMPDOR-----SCTDIDO--CEDNPNICDGGQCTN 1256
 QY 130 -PGTYGPDCLAOG-----GSORPCSGNGHC-----SGD-GSRQDGSRCRHMGYGP 175
 DB 1257 IGEY--RCLCYDGFMASSDMKTCVDVNECDLNPNLCSLTGCEENKRGSRICHCMDKYSK 1314
 QY 176 L-----CTDCMDYFSSLNETHSICTACDESKTCSGLTNRDCGCEVGMVLDEGACVDV 231
 DB 1315 KKGTCITDINECEIGHNCGRAHVCNTTAGSPK-CS-----CSPGMIDGJICTDL 1364
 QY 232 DECAEPPPCSAQPCNANGSYTC-----BEC-----DSSCYGT 267
 DB 1365 DECSNGTHMSCSHADCKNTMGYSRLCKDGYTGDEFTCTDIECSBNLMLCGNGCQLNAP 1424
 QY 268 G-----EGP-----GNCK-----ECISGYA--REHQ 287
 DB 1425 GGYRCECDMGFPVSADGKACEDIDECSLPNICVFCHNLPLFLRCCEIGYELDRSGN 1484
 QY 288 CADVDECSLAERTCYARKNENCYNTPGSYVCVPDGFEE--ETEDACV 331
 DB 1485 CTDVNEC-LDPTTCT--SGNCVNTPGSYTCDCSPDELPNTRVGCV 1527

RESULT 13

A57278
 fibrillin-2 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 11-Jan-2000
 C:Accession: A57278
 R:Zhang, H.; Hu, W.; Ramirez, F. J. Cell Biol. 129, 1165-1176, 1995
 A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extrac
 A:Reference number: A57278; MUID:95263670
 A:Accession: A57278
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2907 <ZHA>
 A:Cross-references: GB:L39790; NID:9762830; PIDN:AAA74908.1; PID:9762831
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:1239-1274/Domain: EGF homology <EGF>
 F:2488-2523/Domain: EGF homology <EGF>

Query Match 12.6%; Score 252.5; DB 2; Length 2907;
 Best Local Similarity 24.8%; Pred. No. 1.6e-08;
 Matches 103; Conservative 30; Mismatches 123; Indels 159; Gaps 25;

QY 23 EAAKRPTE-HRCRGLVDKFNQGMWDYAKNNGGNTAMEEKTLSKYESSEIRLEILEG 81
 DB 2249 ECAQNPLCAFRCMNTGSEY-----CTCPVGYG-----LREDKMKC-----DLDECAEG 2294
 QY 82 L--CESSDEPCOMLEAOEHLLEAMWLKSEYPDLEFMEVCYKTKVCSPPG-TYGPDL 138
 DB 2295 LMDCEGRGMCKNL-----IGTFMCICPGRMARPPDE 2327
 QY 139 ACQGGSQ--RP--CSGNGHSGDGSRODGSRCRHMGYGPGLC--TDCMDG----- 183
 DB 2328 GCVDENECHTRKPGICE-NRCV--NIISYRCECNEBFGQSSSGTECLDNROGLCPAFV 2383
 QY 184 -----YFSSLN-ETHSICTACD----- 200
 DB 2384 LQTMQMASSSRNLVTKSEB--CCDGRGWHQCELOPLPRTAQYKRIKCPHGGYATDGRD 2442
 QY 201 -ESCKTCGL-TNROGE-----CEVGMVLDEG--ACVVDCAEAPPCSAQPFCK 248
 DB 2443 IDECKVMSLCTNQCVTMTGSGFRCPCKVGYTMDISGTRACVDLDECSOSPKPCNF--ICK 2500

OY 249 NANGSYTC-----ECDS-----CVGCTGEGPNCCKEISGYAR 283
 Db 2501 NTKGSYQSCSPRGVVLGDEGDKTKDDECKTKQKHNCOFLCNLTIG---GFYCKKCPPEFTQ 2557
 OY 284 EHGOCAVDDECSLAETKCYKKNENCTYTPGSYVCVCPDGE-----ETEDAC 330
 Db 2558 HHNACIDNNBEGSGPSLCGAKG-ICQNTPGSFCECGRFLDASGLNCEVDDEC 2611

RESULT 14

T09456
 Intrinsic factor-B12 receptor Cubillin precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 A:Accession: T09456
 R:Kozuyaki, R.; Kristiansen, M.; Silhataroglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N.
 Blood 91, 3593-3600, 1998
 A>Title: The human intrinsic factor-vitamin B12 receptor, cubillin: Molecular characterization.
 A:Reference number: 216677; MUID:98241400
 A:Accession: T09456
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3623 <KOZ>
 A:Cross-references: EMBL:AF034611; NID:93929528; PIDN:AC82612.1; PID:93929529
 C:Genetics:
 A:Map position: 10p12
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: receptor; vitamin B12 uptake
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <MAT>
 F:436-467/Domain: EGF homology <EGF>

Query Match 12.58; Score 250; DB 2; Length 3623;
 Best Local Similarity 29.18; Pred. No. 2.8e-08;
 Matches 76; Conservative 26; Mismatches 95; Indels 64; Gaps 16;
 OY 118 WFCYKTLKVCSPCTYPDCLA-----CG--GSGORPCSGNCHGSDSGSCRGCHMG 171
 Db 154 FFLC-----CPQWKPPPLCSADVNECEIYSGTFLSCONGTCV---NTMKSYSCHCPPE 204
 OY 172 YOGFLCT-----DCMDGYFSSLNETHSICACDESCKTCGTLNRDGECEVAVL--DE 225
 Db 205 TYGPGQASKYDDECG--SVARCVHGI---CEDLMREQAGEPRYSC-VCDAGMSSNS 257
 OY 226 GAC-VVDDECAAEPPPSAAQFCNNANGSYTCCECDSSCVGCTGEGPNCCKEISGYARE 284
 Db 258 PACTLDHDECSFPGPCSTLVOCFNTGSPFCGACPTGMOG-----NGYI-- 302
 OY 285 HGOCADVDECSLAETKCYKKNENCTYTPGSYVC-VCPDGE-----ETEDAC 329
 Db 303 ----CEDINECEINNGGCSVAPPEVCVNTPGSSHCQACPGTGGGRVCTLTIDICSVSNGG 359
 OY 330 CVPPEAEATEGESPP--TOLP 348
 Db 360 CHPDASSSTLGLSLPCTCLP 380

RESULT 15

A54105
 fibrillin-2 precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 17-Nov-2000
 A:Accession: A54105; S17063; S31101
 R:Zhang, H.; Apfelroth, S.D.; Hu, W.; Davis, E.C.; Sanguinetti, C.; Bonadio, J.; Mecham, J.; Cell Biol. 124, 855-863, 1994
 A>Title: Structure and expression of fibrillin-2, a novel microfibrillar component preferentially expressed in the aorta
 A:Reference number: A54105; MUID:94165150
 A:Accession: A54105
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-2918 <ZHA>

A:Cross-references: GB:U03272
 R:Lee, B.; Godfrey, M.; Vitale, E.; Horl, H.; Mattei, M.G.; Sarifrazi, M.; Tajpour, N.
 Nature 352, 330-334, 1991
 A>Title: Linkage of Marfan syndrome and a phenotypically related disorder to two diff
 A:Reference number: S17062; MUID:91304567
 A:Accession: S17063
 A:Molecule type: mRNA
 A:Residues: 752-1489, 1791, 'GS', 1794-1796, 'QLI', 1922-1923, 'LD', 1926, 'P', 1928 <LEB>
 A:Cross-references: EMBL:X62009

R:Milwicz, D.M.
 Submitted to the EMBL Data Library, December 1992
 A:Reference number: S31101
 A:Accession: S31101
 A:Molecule type: mRNA
 A:Residues: 752-1407, 'R', 1409-1489, 1791, 'GS', 1794-1796, 'QLI', 1922-1923, 'LD', 1926, 'P',
 A:Cross-references: EMBL:X62009
 C:Genetics:
 A:Gene: GDB:FBN2
 A:Cross-references: GDB:128122; OMIM:121050

A:Map position: 5q23-5q31
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: extracellular protein
 F:1-29/Domain: signal sequence #status predicted <SIG>
 F:30-2918/Product: fibrillin-2 #status predicted <MAT>
 F:1245-1280/Domain: EGF homology <EGF>
 F:1970-2013/Domain: EGF homology <EGF>

Query Match 12.48; Score 249.5; DB 2; Length 2918;
 Best Local Similarity 21.18; Pred. No. 2.6e-08;
 Matches 100; Conservative 53; Mismatches 125; Indels 197; Gaps 26;

OY 33 RCRLVDFK-----NQGM-VDTAKN-----FGGG---NT--AAEETLSKYE 69
 Db 1087 KCRNTISFKCRNSGALDMEERNCTDIDECRISPDLCGSGICVNTPGSFCECFEGYE 1146
 OY 70 SSEIRLEILE-----GLCESSD--FEC-----NOMLEAOEHLIEMWLOLK 109
 Db 1147 SGFMAMKNCMIIDCERNPPLRCGTVCVNTGSGFQCCPLHELSPSDECDINEGLS 1206
 OY 110 SEYDPLEFWFC-----VTKLVCCSPGTYG--PD-----CLACGGSGORPCSGN 151
 Db 1207 DN-----LCRNKCVMIIGTYQSCNPGYQATPDRGCTYDIDECMIMNGCPTQCTNS 1259
 OY 152 GHCGSDSGRODSCRCRMGY-----GQPLCTD-----CM 181
 Db 1260 -----EGSYECSSSEGYALMPDRSCADIDECENNPIDCGGQCTNIPGEYRCICY 1310
 OY 182 DGYFSSLNETHSICACDESCKTC-----SGLTN--RDGCEC 217
 Db 1311 DGFAMSHMKICIDVNECDLNSNIMCECENTGSPFICHOGLGYVKKGTTCGDVDEC 1370
 OY 218 EV-----GWLDEGACVDVDECAAEPPPSAAQFCNNANGS 253
 Db 1371 EIGHANCDMHASCLNIPGSFSCREGWIGIKICIDIDECSSNGTHOCSINACVNTPGS 1430
 OY 254 YTCEDSSSCVGTGEG-----PGNCK-ICISGY--AREHG 286
 Db 1431 YRC-ACSE--GFTGDGTCTSDVDECAENILNCEGCLNPGAVRCBCEGFTPASDR 1486
 OY 287 QCADVDECSLAETKCYKKNENCTYTPGSYVCVCPDGEFEF-----DACVPP 333
 Db 1487 SCQIDDECSF-QNICY--SGTCNNLPGMFHICIDCDGYLDTGTGNCIDIDECADP 1538

RESULT 16

A48225
 subtilisin-like proprotein convertase (PC 3.4.21.-) PC5 precursor - mouse
 N:Alternate names: kexin homology; serine proteinase PC6
 C:Species: Mus musculus (house mouse)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: A48225; JX0248
 R:Lusson, J.; Vieau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.

Query Match	12.4%	Score 248	DB 1	Length 915
Best Local Similarity	30.0%	Pred. No. 1.3e+08		
Matches	66	Conservative	20	Mismatches 86
			Indels	48
			Gaps	13

```
Qy      297 A-----EKTVCVRNENCY--MTEGSIYC-VCPDGF 323
          : | | : : | | | | | | :
Db      823 SYILDHSEGGYISCKRCDNSCLTNGGEFGKNCSSCPSEY 862
```

A;molecule type: mRNA

F;26-85/Domain: Gla domain homology <GLA>

```
F;42-6/5/Product: plasma protein S #status predicted <MAT>
F;88-116/Domain: thrombin-sensitive #status predicted <THS>
```

F;161-199/Domain: EGF homology <EG2>
F;205-241/Domain: EGF homology <EG3>

Query Match	12.3%	Score 247;	DB 1;	Length 675;
Best Local Similarity	26.9%	Pred. No. 1.2e-08;		
Matches 93;	Conservative 34;	Mismatches 113;	Indels 106;	Gaps 21.

```

OY      287  --QCADVDECSLAECTVRKNENCYNTPGSIYVCV--PDGEETED 328
          | | | | | | | | | | : | | | | | | | | | | : |
Db      238  SKSCKDVDECS--ENTCA--QICLVNYPGGISCYCDGKKGFKLAD 278

```

RESULT 18

A;molecule type: mRNA

A;molecule type: protein

A:Residues: 236-238, 'X',240-247,260-275,336-344, 'L',346-361,405-426,566-568, 'PM',569
C:Superfamily: unassigned EGF-related proteins; EGF homology
G:Keywords: calcium binding; duplication; extracellular matrix; glycoprotein; homotrimer
F:942-978/Domain: EGF homology <EGF>

Query Match	12.3%	Score 247	DB 2	length 1222
Best Local Similarity	22.1%	Pred. No. 1.9e-08		
Matches 94	Conservative 30	Mismatches 132	Indels 170	Gaps 17

```

QY      83 CESS-----DFFCNOML-----EAOEHLBAWMLQJLKS 110
      III:      : II:      : II:      :

```


OY 270 ---GPGNCKECSIGYAREHQCADVDEC-----SLAEKTCVAKNENCNTPGSYVCVP 320
 DB 270 NTLGTYGDSCTGTGSGDGNVCVKDSCVKNKCHKLA--TC-KVYDDGTSNAGDITCYCP 326
 OY 321 DGFEEET---EDACVPPA 334
 DB 327 DGVYDGDIGIEGECVKS 343

RESULT 21

A55184
 fibulin-2 precursor - human
 N:Alternate names: protein DKFP586A1519.1
 C:Species: Homo sapiens (man)
 C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000
 C:Accession: A55184; T08744
 R:Zhang, R.Z.; Pan, T.C.; Zhang, Z.Y.; Mattei, M.G.; Timpl, R.; Chu, M.L.
 Genomex 22, 425-430, 1994
 A:Title: Fibulin-2 (FBLN2): human cDNA sequence, mRNA expression, and mapping of the gene
 A:Reference number: A55184; MUID:95104855
 A:Accession: A55184
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1184 <ZHA>
 A:Cross-references: GB:X82494; NID:9575232; PIDN:CAA57876.1; PID:9575233
 R:Mambuti, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: 216471
 A:Accession: T08744
 A:Molecule type: mRNA
 A:Residues: 656-719, /ODECLMGANDCSRRQFCVNTLGSFYCVNHFTVLCADGYTLNHRKQVD, /720-853, /T, /855-11
 A:Cross-references: EMBL:AL050095
 A:Experimental source: adult uterus; clone DKFP586A1519
 C:Genetics:
 A:Gene: GDB:FBLN2
 A:Cross-references: GDB:293037; OMIM:135821
 A:Map position: 3p25-3p24
 A>Note: DKFP586A1519.1
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: alternative splicing; extracellular matrix
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-1184/Product: fibulin-2 protein #status predicted <MAT>
 F:905-941/Domain: EGF homology <EGF>

Query Match 12.1%; Score 242.5; DB 2: Length 1184;
 Best Local Similarity 23.8%; Pred. No. 3.5e-08;
 Matches 88; Conservative 38; Mismatches 115; Indels 129; Gaps 19;

OY 83 CESS---DEECNOML-----EAOEHLBAMWLQKLS 110
 DB 535 CESSNPLGYPCNHHVMSCEGEPLIVPRRPREPAAPRRVSEEMAGRELISGTA 594
 OY 111 EYP-----DLEFWFCVKT---KYCCSPG-TYGPDCILAC--QGSQRP- 147
 DB 595 ELPNLSLPDGDDECLLPEGLCOHLICINTVGSYHCACFPFESLDDGRYCRPGHPPE 654
 OY 148 -----CSNGHCSGDSROGDG--CGCHNGY-----QG 174
 DB 655 AQPAPAKSEESQAVASNTPLPLPOPNTCKDNGPKQVCSYVGGSAICSPFGYALIMADG 714
 OY 175 PLCTD---CMDYFSLRNETHSICACDESCRSGLTNRDCGECEVGVLDGACVY 231
 DB 715 VSCEDINECVTLHLHCSRRE-HCVNLTGSHFC--YKALF-----CEPPTALKDGECEV 765
 OY 233 DECAAPPPCSAAQPCCKNANGSYTCEECDCSCVCGTGEPPGCKECISYABE-HQCAD 290
 DB 766 DECAAGTHTCOPGFLCQNKSGSFYCOA-----RORCMDFLDDPENGVD 810
 OY 291 VDECSLAETCVKKNENCNTPGSYVC-----VCPDGFEEETDA--CVPAAAE-----A 338
 DB 811 INECTSLSEPC-RPGFSCLINTVGSYTCQRPNLICARGHASDDGAKCVVNECEGTGVHRC 869

OY 339 TEGESPTOLP 348
 DB 870 GEGQYCHNLP 879

RESULT 22

A35626
 transforming growth factor beta-1-binding protein - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Sep-1990 #sequence_revision 21-Sep-1990 #text_change 11-Jan-2000
 C:Accession: A35626
 R:Kanzaki, T.; Olofsson, A.; Moren, A.; Wernstedt, C.; Hellman, U.; Miyazono, K.; Cla
 Cell 61, 1051-1061, 1990
 A:Title: TGF-beta1 binding protein: a component of the large latent complex of TGF-beta
 A:Reference number: A35626; MUID:90275601
 A:Accession: A35626
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1394 <KAN>
 A:Cross-references: GB:M34057; NID:9339547; PIDN:AAA61160.1; PID:9339548
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: alternative splicing
 F:750-791/Domain: EGF homology <EGF>

Query Match 12.0%; Score 240.5; DB 2: Length 1394;
 Best Local Similarity 27.1%; Pred. No. 5.4e-08;
 Matches 85; Conservative 22; Mismatches 96; Indels 111; Gaps 19;

OY 67 KYSESEIRLETLLEGLCESSDFECNOMLEAOEHLBAMWLQKSEYPDLEFWFCVTKLV 126
 DB 655 EYCDGGRMTQ--RGRCEDID-ECLNPSTCPDEQ-----CVN----- 688
 OY 127 CCSPTYGPDLACCGGSGRSPGSGNHC-----SGDGR-QGDSGCRCHNGY 172
 DB 689 --SPGSY-QCVPTGEFF--GWNQCCLDVDECLPYNVCAAGDSSNLEGSYMCCHNGY 741
 OY 173 QGPLCTDCMDGYFSLRNETHSICTACDE-----SCYTCSGLTNRDCGECEVGV 222
 DB 742 -----TRPDHKNCHDIDECQCGGNLCVNGGCKNTGSPRCTGQ---GYQ 783
 OY 223 LD--EGACVDDECAAPPPCSAAQPCCKNANGSYTCEBCDSSCVGCTGBGPNCKECSIG 280
 DB 784 ISAAKDCEIDIDEC-QHRLCAHQ-CRNTGESSFOC-VCDQ---GYRASGLD----- 830
 OY 281 YAREHQCADVDECSLAETCVKKNENCNTPGSYVCVCPDGE-----ETE 327
 DB 831 -----HCEIDINECELDKSVCOR--GDCINTAGSYDCTCPDGFQLDNKTCCODINECEHP 882
 OY 328 DACVPPAEATEEG 341
 DB 883 GLCGPQGECLNTEG 896

RESULT 23

KXHU5

plasma protein S precursor - human
 N:Alternate names: vitamin K-dependent protein S

C:Species: Homo sapiens (man)
 C:Date: 21-Sep-1990 #sequence_revision 26-Jan-1996 #text_change 16-Jul-1999
 C:Accession: A35610; A35611; A26157; A25891; A35612; A60903; S02424; S09519
 R:Schmidel, D.R.; Tatro, A.V.; Phelps, L.G.; Tomczak, J.A.; Long, G.L.
 Biochemistry 29, 7845-7852, 1990
 A:Title: Organization of the human protein S genes.
 A:Reference number: A35610; MUID:91084444

A:Accession: A35610
 A:Molecule type: DNA
 A:Residues: 1-676 <SCH>
 A:Cross-references: GB:M57853; NID:9190547; PIDN:AAA60357.1; PID:9190549; GB:702917
 A>Note: the authors translated the codon TTT for residue 26 as Leu
 R:Piets van Amstel, H.K.; Reijtsma, P.H.; van der Logt, C.P.E.; Bertina, R.M.
 Biochemistry 29, 7853-7861, 1990

Db 814 GRC--VQSCSVSYLDHSLGEGYKCKRCNDNSCLTNGPFRKNCSCPSGY 862

RESULT 25

743251
 furin (EC 3.4.21.75) - fall armyworm
 N:Alternate names: paired basic amino acid cleaving enzyme; proprotein convertase; serin
 C:Species: Spodoptera frugiperda (fall armyworm)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
 C:Accession: T43251
 R:Cieplik, M.; Klenk, H.
 Submitted to the EMBL Data Library, January 1996
 A:Description: Cloning and functional characterization of furin from Spodoptera frugiper
 A:Reference number: 22368
 A:Accession: T43251
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1299 <CIE>
 A:Cross-references: EMBL:Z68888; NID:q1167859; PID:q219690; PIDN:CAA93116.1
 A:Experimental source: clone Sfurin 6; ovary
 C:Function:
 A:Description: responsible for the endoproteolytic processing of proproteins with speci
 C:Keywords: hydrolase; serine proteinase

Query Match 11.9%; Score 239; DB 2; Length 1299;

Best Local Similarity 25.8%; Pred. No. 6.4e-08;
 Matches 92; Conservative 32; Mismatches 146; Indels 86; Gaps 20;

QY 19 PPAPEAKKPPCHRCGLVDFKNOGMVDYAKKNGGNTAMEKTL--KYE--SSEIR 74
 DB 881 PPAAYADKKRKKCKMCP-----VGCTCTSAFCLSCPEKMKLKKKG 922
 QY 75 LLEILEGCESSDF-----ECNOMLEA---QEHLAMWLQKSEYPDF-EMFCVK 122
 DB 923 CMPVSGDKCSAGEFAVDQKCRKNPACDSYGENEGHC-----LTCPPNLLQDKCVP 976
 QY 123 TLKVCSPPTYGP--DCLACOGSGORPCSGNGHSGDGRSGDCRCHMGYQGPLC-TD 179
 DB 977 E-----CSKRYVAEGACARCAMHG-----CSDCVRLNCTSCASLRLRQSGCRIS 1022
 QY 180 CNDGFSSLRNETHSICACDSCSKTSGSLTNRDGCECEVGVLDGACVDYDECAEPP 239
 DB 1023 CADGYRAD-----RGTCCKYLSICRCLIPRRDQCAACDEGRMLAAGC-----HP 1068
 QY 240 PCSAOFPCNANGSYTCECDSSCVGCTGEGPNCCKECSGAREHGOCAD-----VDE 293
 DB 1069 ECPQG-FYQSPG--CRHCHHYCRECDGSGPLHCKSCPFRFMDGLCECLGSOYDA 1124
 QY 294 CSLAERTCVKKNKNCYNPFGSYVC--VC--PDGFEETEDACVP-PAPAEATEGESPT 345
 DB 1125 TSGTCSGSCDASGRTG--SGRGQSGCTTCSRLRIDRLNNCVCPCCSRGVTNSTPT 1179

RESULT 26

602428
 subtilisin-like proprotein convertase (EC 3.4.21.-) 5 precursor - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Mar-2000
 C:Accession: G02428
 R:Reudelhuber, T.L.
 Submitted to the EMBL Data Library, February 1996
 A:Reference number: H01242
 A:Accession: G02428
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-899 <RED>
 A:Cross-references: EMBL:U49114; NID:q1218057; PIDN:AAA91807.1; PID:q1218058
 C:Genetics:
 A:Gene: PCS
 C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C:Keywords: hydrolase; serine proteinase

F:148-386/Domain: subtilisin homology <SBT>

Query Match 11.8%; Score 236; DB 2; Length 899;

Best Local Similarity 25.8%; Pred. No. 7.4e-08;
 Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKFNQGVDTAKKNGGNTAMEKTLKYSSEIRLEILEGCESSDFECNOMLEAQ 97
 DB 598 VERFRRSREDPTDYGTEADYGP-----CDPECSYV-----GDDGPGPHCNCLH-- 644
 QY 98 EHLAMWLQKSEYPDFEMFCVTKL-----KYC--CSP-----GTYGPDCLAC 140
 DB 645 -----YYTKLKN-----TRICVSSCPGHYHADKRRKCAKPCSCFSGHGQCMSC 693
 QY 141 QCG-----SQRPCSGNGHSGDGRSG--DQSCR-----CHMGVQGLCTDQMDQYSSLRN 190
 DB 694 KYGYFLNEETNSCVTHCP--DGSYQDTKKNLCKKSENCCKTCTEHNCTEGRGL--SLQG 750
 QY 191 ETHSI-----CTACDESKTCSGLTNRDGCECEVGVLDGACVDYDECAEPP 238
 DB 751 SRCSVSCDEGRFRFNGODQPCHRFCATCAGAGADGCINCITEGYFMEDGRVQ----- 802
 QY 239 PCSAAGFCFKA--NGSYTCECDSSCVGCTGEGPNCCKECSGYAREHGO-----CAD 290
 DB 803 -SCSISYFDHSSENGYSKCKCDISCLTNGPFRKNCSPSGYLDLGMCOMGAICD 861
 QY 291 VDECSLAEK---TCYKKNKNCYNTPGSYVC 317
 DB 862 ATEESABSGFCMLVKKNNLCKRKVLOOLC 891

RESULT 27

subtilisin-like proprotein convertase (EC 3.4.21.-) homolog - human
 N:Alternate names: PCE4 protease
 C:Species: Homo sapiens (man)
 C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 22-Jun-1999
 C:Accession: J06148
 R:Miranda, L.; Wolf, J.; Pichuanes, S.; Duke, R.; Franzusoff, A.
 Proc. Natl. Acad. Sci. U.S.A. 93, 7695-7700, 1996
 A>Title: Isolation of the human PC6 gene encoding the putative host protease for HIV-
 A:Reference number: J06148; MUID:96353880
 A:Contents: CEM T-cell
 A:Accession: J06148
 A:Molecule type: mRNA
 A:Residues: 1-915 <MIR>
 A:Cross-references: GB:U56387; NID:q1498312; PIDN:AAC50643.1; PID:q1498313
 C:Comment: This protein functions as a soluble enzyme within the secretory pathway. I
 C:Genetics:
 A:Gene: pce4
 C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C:Keywords: glycoprotein; hydrolase; serine proteinase
 F:164-402/Domain: subtilisin homology <SBT>

Query Match 11.8%; Score 236; DB 2; Length 915;

Best Local Similarity 25.8%; Pred. No. 7.5e-08;
 Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKFNQGVDTAKKNGGNTAMEKTLKYSSEIRLEILEGCESSDFECNOMLEAQ 97
 DB 614 VERFRRSREDPTDYGTEADYGP-----CDPECSYV-----GDDGPGPHCNCLH-- 660
 QY 98 EHLAMWLQKSEYPDFEMFCVTKL-----KYC--CSP-----GTYGPDCLAC 140
 DB 661 -----YYTKLKN-----TRICVSSCPGHYHADKRRKCAKPCSCFSGHGQCMSC 709
 QY 141 QCG-----SQRPCSGNGHSGDGRSG--DQSCR-----CHMGVQGLCTDQMDQYSSLRN 190
 DB 710 KYGYFLNEETNSCVTHCP--DGSYQDTKKNLCKKSENCCKTCTEHNCTEGRGL--SLQG 766
 QY 191 ETHSI-----CTACDESKTCSGLTNRDGCECEVGVLDGACVDYDECAEPP 238

Db 767 SRCVSCEDGRYFNGCQCCHRECATCAGAGADGCINTGEGYMEDGRVQ----- 818
QY 239 PPSAAGFCMNA--NGSYTEECDDSSCVGCTGEGPNCCKECISYAREHQ-----CAD 290
Db 819 -SCSISYFPHSSSENGKSKCKDISCLTNGPFGKNCSTCPSCGYLLDLGCOMGAICKD 877
QY 291 VDECSLAEK---TCVRKNENCYMTPGSYVC 317
Db 878 ATEESMAEGGFCMLVKKNNLCQKRVLQQLC 907

RESULT 28
T18739
hypothetical protein B0393.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18739
R:Stulston, J.
Submitted to the EMBL Data Library, September 1994
A:Reference number: Z19013
A:Accession: T18739
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1106 <MIL>
A:Cross-references: EMBL:Z17983; NID:e1519039; PIDN:CAA86058.1; GSPDB:GN00021; CESP:B0393
A:Experimental source: clone B0393
C:Genetics:
A:Gene: CESP:B0393.5
A:Map position: 3
A:Introns: 15/3; 60/1; 100/3; 343/2; 443/1; 507/3; 609/3; 659/1; 796/1; 876/1; 946/1; 99

Query Match 11.7% Score 234; DB 2; Length 1106;
Best Local Similarity 22.2% Pred. No. 1.2e-07;
Matches 84; Conservative 26; Mismatches 100; Indels 168; Gaps 14;

QY 108 LKSEYPLD---FEW-FCVTKLV-CCSPGYGPDCLACGGGSG--RPGSGNGHC-SGDGS 159
Db 156 LLSKSDIGIGEWLFRVDDDRITLCGAGFKGLEICSCASQWPNDCSKSCHDGDGDC 215
QY 160 RQGDGSC--RCHMGYOG-PLCTDCCMD-----GYFSSL----- 188
Db 216 DOENGRCPNKCSGPGWIGEBICDEMDCEMGIDNCPEQPDCLNTPGSFLCLCFEYDEA 275
QY 189 -----RNEHHS----- 194
Db 276 QQCKNSKAAPSPAPIDVVI PMHPTFRKPTPPKVTAPPRNRPKSTTAATSPFITETT 335
QY 195 -----ICTACDESCKTCGSLTNRDGCE 218
Db 336 TRQTTTTRKESITPLLTTEVYTRPSPTTYVTAPACARCDQAKNSNGVCT-----CS 390
QY 219 VGVNLDEGACVDVDECAEPSPCSAOPCKNANGSYTCEECDDSSCVGCTGEGPNCCKECI 278
Db 391 EGFGLDGFRCYDVECEIPAVCRDHSICSWTIGSFEC-----TCH 431
QY 279 SGVAREHGACADVDECSLAKTC--VAKNENCYMTPGSYVCVCPDGE----- 324
Db 432 GGRFEEGKCEKEDVECELPKICGDPNKGTKLCKDKGTFFCLCKDGYEGDPSSECRDVE 491
QY 325 -ETEDACVPAAEAATEG 341
Db 492 CKNDPACGPNSSQCTNTGG 509

RESULT 29
KXMS
plasma protein S precursor - mouse
N:Alternate names: vitamin K-dependent glycoprotein S
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 26-Jan-1996 #text_change 16-Jul-1999
C:Accession: S43504; I59616; S35962

R:Chu, M.D.; Sun, J.; Bird, P.
Biochim. Biophys. Acta 1217, 325-328, 1994
A:Title: Cloning and sequencing of a cDNA encoding the murine vitamin K-dependent pro
A:Reference number: S43504; MUID:94198297
A:Accession: S43504
A:Molecule type: mRNA
A:Residues: 1-675 <CHU>
A:Cross-references: EMBL:Z25469; NID:g396426; PIDN:CAA80961.1; PID:g396427
R:Lu, D.; Schmidt, D.K.; Long, G.L.
Thromb. Res. 74, 135-142, 1994
A:Title: Structure of mouse protein S as determined by PCR amplification and DNA sequ
A:Reference number: I59616; MUID:94302659
A:Accession: I59616
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 35-492, V, 494-675 <RES>
A:Cross-references: GB:I27439; NID:g487866; PIDN:AAA40006.1; PID:g487867
C:Complex: in plasma forms a complex with C4b binding protein
C:Function:
A:Description: a cofactor for activated protein C (EC 3.4.21.69); thrombin cleavage p
C:Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat
C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; blood coagulation; car
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-41/Domain: propeptide #status predicted <PRO>
F:42-675/Product: plasma protein S #status predicted <MAT>
F:121-154/Domain: EGF homology <EG1>
F:161-199/Domain: EGF homology <EG2>
F:205-241/Domain: EGF homology <EG3>
F:247-282/Domain: EGF homology <EG4>
F:315-666/Domain: sex hormone-binding globulin homology <SHB>
F:325-478/Domain: laminin G repeat homology <LGR>
F:47-48, 55-57, 60-61, 66-67, 70-73, 77/Modified site: gamma-carboxyglutamic acid (Glu) #
F:58-63, 88-113, 121-134, 126-143, 145-154, 161-175, 171-184, 186-199, 205-217, 212-226, 228-24
F:111-112/Cleavage site: Arg-Ser (thrombin) #status predicted
F:126/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F:177, 219, 258/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:499, 509/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.6% Score 232; DB 1; Length 675;
Best Local Similarity 26.3% Pred. No. 1.1e-07;
Matches 91; Conservative 30; Mismatches 119; Indels 106; Gaps 20;

QY 11 LRLPLLLPAPPAEAKKPTPCHRCGLVDKRNQGMVDTAKNFGCGNTAMEKLTSTYES 70
Db 11 LLAFLAVTPVSE-----NPLSKERASQVIVRRR---ANTLFETKGNLER 56
QY 71 SEIRLEITLEGICSSDFECNQMLEAOEH-----LEAWML-----OLKSEYP 113
Db 57 -----ECIEELCKKE--EAEEVEENNPETDYFPKYLGCIGAFRVGSFHAARQSANAYP 108
QY 114 DLFEWFCVTKLVCCSPGYTPD--CLACQGGSGRPGSGNGHSGSDGSRQDGSRCRCHMY 172
Db 109 DLRS--CVKATISDCCDIPICNEDGYLACQDS-----QAAFTCFCKPGW 149
QY 173 QGPLC-----IDCMGIFSSILANETHSICTACDESCKTCGSLTNRDGCECEGWL--DEG 226
Db 150 QGDRGQYDVNCKD-----PSNVNGGCSQICDNTPG--SYHC--SCKRGFAPMLPKK 197
QY 227 ACVVDDECAEAPPSPCSAOPCKNANGSYTCEECDDSSCVGCTGEGPNCCKECISYAREHG 286
Db 198 DCKLDECALKPSVCGTA-VCKNIPGDEEC-----ECPPDGYRRDPS 237
QY 287 --QCADVDECSLAEKTCVRKNENCYMTPGSYVC--PDGFEETED 328
Db 238 SKSCKDVECS--ENMCA---QLCVNFPGGYSCYCDGKKGFKLAOD 278

RESULT 30
G00043
osteonidogen - human
C:Species: Homo sapiens (man)

```
C>Date: 13-Mar1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: G00043
R:Ohno, I.
submitted to GenBank, July 1996
A:Reference number: H00043
A:Accession: G00043
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-1376 <DN>
A:Cross-references: GB:086425; NID:g1449166; PIDN:BAAL3087.1; PID:g1449167
C:Superfamily: nitrogen; EGF homology; LDL receptor YWTD-containing repeat homology; thy
F:806-843/Domain: EGF homology <EGF1>
F:853-891/Domain: EGF homology <EGF1>
F:941-1006/Domain: thyroglobulin type I repeat homology <THY1>
F:1020-1085/Domain: thyroglobulin type I repeat homology <THY2>
F:1242-1286/Domain: LDL receptor YWTD-containing repeat homology <YW3>
```

Query Match	11.6%	Score 233;	DB 2;	Length 1376;
Best Local Similarity	27.8%	Pred. No. 1.9e-07;		
Matches 63;	Conservative 12;	Mismatches 74;	Indels 78;	Gaps 7;

QY	147	PCSNMGHSGGSGNGGSGSQRCHMGYGGRPLCTDOMDGFSSLNENHSHIACDECKTC	206
		: :	
Db	763	PCY-----DGHMCCTTACBPG-----TGVDYTC--	787
QY	207	SGLTRNDCGEGECVWVLDEACAVDVECAAEPPCSAAQCFKNANGSYTCE-----	257
		: : :	
Db	788	-----EASVSYGGDGNVCNDENCAIGFHRCGNSVCINLPSSVRECESGSYEPAD	838
QY	258	-----ECDSGVCNCTGEGFGNCKE-----CISGVAAREHGQADVDECS	295
		: :	
Db	839	DRHFCILITPPANCPCEGDSHTCAAPAGCARVHHGSGSTFSCACIPGYAGDGHQCTDVECS	898
QY	296	LAETCYRKNNKNCYNTGSGYVCVPDSEETEDACVPAALAEATGE	342
		: :	
Db	899	---ENRRC--HPAATCYNTGSGESCRCRQPPYGYDGFQCLITDSTSSLTPE	942

RESULT 31
T42681
hypospectral protein DKFzp434E0321.1 - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T42681
R:Blum, H.; Bauesachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A:Reference number: Z22233
A:Accession: T42681
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-870:871-1069 <MA>
A:Cross-references: EMBL:AL133021
A:Experimental source: adult testis; clone DKFzp434E0321
A:Note: the cDNA sequence contains a +1 frameshift near codon 870
C:Genetics:
.:Note: DKFzp434E0321.1

	Query Match	11.5%	Score 231.5;	DB 2;	Length 1069;
	Best Local Similarity	26.7%;	Pred. No. 1.6e-07;		
Matches	66;	Conservative	23;	Mismatches 107;	Indels 51; Gaps 11.
OY	127 CCSPTVGPDLACGGGSGRPCSNGHCSGDSSRCRGDSCRHMVGYPGLCTCDMDGYRS	186			
	: : : : : : :	:	: : : : : :	:	:
Db	504 CCK-GYFRDCQACGGADAPCNRNRYCLDOYSATGE--CKNTGFNGTACMCMPGRAG	560			
OY	187 SLRNETHSIICNADESKTCISGLTRNDGEGCEVGVNLDEBGACVDVDEC--AAEPPCSA	244			
Db	561 -----PDLPGCGSDHGCCDDGITGSGCCLETGTWT--GPSCTQAVMLPAVCTPPCSAH	612			
OY	245 QFCNKANGSYICE---EDDSVCVGT-----GEIGNCKE-----CLSTGA	282			
	: : : :	:	: : : : :	:	:

```

Db 613 ATCKENN---TCQCNILYEGGIGTCFVDDCKDNQGCACVAVCSOKGTFVSCSGCKGYK 669
QY 283 RHHQCAQDVECSLAERTCYVRKNKNCYNT--PGSVVCVCPDGFEEFEDACAPRAEATBGS 341
Db 670 GDGSHCTEIDRCADGLNGCGCHNATCKMTGRGKHKCECKSHYVGDLNC----- 718
QY 342 ESPQLP 348
Db 719 -EPQLP 724

```

```

RESULT      32
S53434
Plasma protein S precursor, vitamin K dependent - rhesus macaque (fragment)
C:Species: Macaca mulatta (Rhesus macaque)
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 16-Jul-1999
C:Accession: S53434
R:Greenard, J.S.; Fernandez, J.A.; Radtke, K.P.; Griffin, J.H.
Biochem J 305, 397-403, 1995
A:Title: Identification of candidate residues for interaction of protein S with C4b b
A:Reference number: S53433; MUID:95134217
A:Accession: S53434
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-642 <GRE>
A:Cross-references: EMBL:L31380
A:Experimental source: tissue type liver
A:Note: the source is designated as rhesus monkey
C:Genetics:
A:Gene: PROS
C:Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat
F:1-51/Domain: Gla domain homology (fragment) <GLA>
F:1-7/Domain: signal sequence (fragment) #status predicted <SIG>
F:8-642/Product: plasma protein S #status predicted <MAT>
F:17-165/Domain: EGF homology <EG1>
F:127-165/Domain: EGF homology <EG2>
F:171-207/Domain: EGF homology <EG3>
F:213-248/Domain: EGF homology <EG4>
F:281-633/Domain: sex hormone-binding globulin homology <SHB>
F:291-444/Domain: laminin G repeat homology <LG>

```

	Query Match	Similarity	11.4%	Score	228.5	DB	2	Length	642;
	Best Local	Similarity	29.0%	Pred.	No. 1.7e-07;				
	Matches	71; Conservative	24; Mismatches	75; Indels	75; Gaps				
Oy	107 QLKSEYPDLEFMFCVVKTLKCCSPETGYPPDCLACGGSSORPCSGNGHCS-GDGSRQGDGS	165							
Db	68 QSTDAYPDLRS--CVNALPDCGSP-----LPCENEGYSCKKG--KASPT	108							
Oy	166 CRCHAGYGGLPC---TPCMDOGFFSLNRETHSICTACDESCCTKSGLTNRCGCEVEWM	221							
Db	109 CTCKRGMGGECEFPIDNECKD-----PSNINGGCSQIOLDNIPG--SYHC-SKSKSF	156							
Oy	222 VL--DEGACVDYDCAEPPPCSAOFCKANANGSTTCECBDSACYGCTGEGFNGKECLS	279							
Db	157 VMLSNKKCKCKVDDECSTLPNNMGTA-VKKNIPGDEEC-----ECPE	196							
Oy	280 GYA--REHGQCADADVDECSLAERTCYRKNNENCYNTPGSIVCVC--PDGEETED-----	328							
Db	197 GYRYLRKLKSKCEDVEDGCS--ENMCA--QLICVNYYGGTYCTCDGRKGFLLAQDNQSCAEAV	251							
Oy	329 -ACVP	332							
Db	252 SVCLP	256							

RESULT 33
A43434
furn (EC 3.4.21.75) 2 - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Mar-2000
C:Accession: A43434

R:Roebroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dunke, U.; Rentrop, M.; Gateff, E.
 J. Biol. Chem. 267, 17208-17215, 1992
 A:Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc
 A:Reference number: A43434, MUID:92381036
 A:Accession: A43434
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1680 <ROE>
 A:Cross-references: GB:M94375; NID:q157461; PID:q157462
 A:Note: sequence extracted from NCBI backbone (NCBIN:111933, NCBI:P.111934)
 C:Genetics:
 A:Gene: FlyBase:Fur2
 A:Cross-references: FlyBase:FBgn0004598
 C:Superfamily: subtilisin homology
 C:Keywords: hydrolase; serine proteinase; transmembrane protein
 F:409-652/Domain: subtilisin homology <SBT>
 F:418-457/638/Active site: Asp, His, Ser #status predicted

Query Match 11.4%; Score 228; DB 2; Length 1680;
 Best Local Similarity 22.8%; Pred. No. 3.9e-07;
 Matches 84; Conservative 31; Mismatches 127; Indels 126; Gaps 19;

OY 31 CHRCGLVDKFNQGVDTAKKNGGNTAWDEKTLTK-----YESSEIRLLETLGLCE 84
 DB 1118 CATCNGPTD---ODCICRASSRY-----AMQNKCLISCPDGFYADKRLKRCMPCQEG-CK 1168
 OY 85 --SSDFECNMLEAEOEHLAMWHLQKSEYRDLFEWFCVTLKVCSPRTYGPCLACQG 142
 DB 1169 TCTISNGVCS-----ECIDNMTLNRKDK-----CIVSGSEGCSESEF---YSQVE 1209
 OY 143 GSORPCSGN-GHCSGDSRQGDSC-----RCHNGYO-----GPLC----- 177
 DB 1210 GQCRPCASGSGCNG-----PADTSCITSPNRLLEQRCVSGCEGFEVFAEGLSLCPLH 1265
 OY 178 -----TDCMDGYFSSLRNETHSITCTACDESCKTCGLTN 211
 DB 1266 TCSQCVSRNCSNCSKGLQLNGECRTTCADGYSD-----RGICACVLSCHTCSGPRR 1320
 OY 212 RDGCECEVGVLDGACVDDECAAEPPPCSAQFCNANGSYCEECDSQVCTEGP 271
 DB 1321 NQCVQCAAGQOLAGEC-----HPECPEGY---KSDFCCQCHHYCKTCNDAGP 1367
 OY 272 GNCKECISGAREHGQCAD-----VDECSLAERTCVKRNENCTNGSYCVCPDGFEE 325
 DB 1368 LACTSCPHSMIDGGLCMELSSGYVDTTATCTCHDSRCSCG-PGQESC----- 1418
 OY 326 TEDACVPP 333
 DB 1419 --KGCVPP 1424

RESULT 34
 A35844
 Xotch protein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 13-Aug-1999
 C:Accession: A35844
 R:Coffman, C.; Harris, W.; Kintner, C.
 Science 249, 1438-1441, 1990
 A:Title: Xotch, the Xenopus homolog of Drosophila notch.
 A:Reference number: A35844; MUID:90385285
 A:Accession: A35844
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-2524 <COF>
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 C:Keywords: transmembrane protein
 F:146-177/Domain: EGF homology <EGX1>
 F:184-215/Domain: EGF homology <EGX1>
 F:222-254/Domain: EGF homology <EGF>
 F:456-487/Domain: EGF homology <EGX2>
 F:757-788/Domain: EGF homology <EGF3>

F:1025-1056/Domain: EGF homology <EGX3>
 F:1924-1956/Domain: ankyrin repeat homology <AN1>
 F:1957-1989/Domain: ankyrin repeat homology <AN2>
 F:1991-2023/Domain: ankyrin repeat homology <AN3>
 F:2024-2056/Domain: ankyrin repeat homology <AN4>
 F:2057-2089/Domain: ankyrin repeat homology <AN5>

Query Match 11.3%; Score 227; DB 2; Length 2524;
 Best Local Similarity 23.9%; Pred. No. 6.2e-07;
 Matches 69; Conservative 20; Mismatches 100; Indels 100; Gaps 14;

OY 126 VC-CSPGTGPDCL-----ACGCGSORPCSGNGHSGSGSRQGDSCRCHMGYGPLCTD- 179
 DB 165 ICKRPPEFHATCKQODINEC---SQNCKNGGQINE---FGSRTCQGNFTGRNDEP 218
 OY 180 -----CMDGYFSSLRNETHSICTA-----CDBSKCTC-----SGLTN 211
 DB 219 YVPCNPSPCLNGCTCROTDTSTYDCTCLPGFSQNCSENIIDCPSNNCRNGTGVGVNT 278
 OY 212 RDGCECEVGVLDGACVDDECAAEPPPCSAQFCNANGSYC-----EECDSS- 262
 DB 279 YNC-QCPPDWT-GQYCTEDVDECOIMPAQNGGCTCHMYGYNVCVNGTGEDCSENI 336
 OY 263 --CVGCTGEGPGNCKE-----CISGYAREHGQ----- 288
 DB 337 DDCANAAHSGATCHDVAASYCEPCPHGRIGLLCHLDNACISNPNCSNCDTNPVNGKA 396
 OY 289 -----ADVDECSLAERTCVKRNENCTNGSYCVCPDGF 323
 DB 397 ICTCPGYTGPGACNNVDDECSLGNPCER-GGRCTNTLGSFQCNCPOGY 444

RESULT 35
 A46019
 Notch-1 protein - mouse
 N:Alternate names: notch protein
 C:Species: Mus musculus (house mouse)
 C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
 C:Accession: A46019; S25144
 R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G
 Genomics 15, 259-264, 1993
 A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog
 A:Reference number: A46019; MUID:93194170
 A:Accession: A46019
 A:Status: not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-2531
 A:Cross-references: GB:211886; GB:S47228; NID:9288502; PIDN:CAA77941.1; PID:9288503
 A:Note: sequence extracted from NCBI backbone (NCBI:P.127318)
 R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenpan, R.J
 submitted to the EMBL Data Library, April 1992
 A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, sugg
 A:Reference number: S25144
 A:Accession: S25144
 A:Molecule type: mRNA
 A:Residues: 1551-2108, 'O', 2110-2114, 'ALP', 2118-2170 <PRA>
 A:Cross-references: EMBL:211886
 C:Genetics:
 A:Gene: notch-1
 A:Map position: 2
 A:Note: proximal region of chromosome 2
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
 F:106-138/Domain: EGF homology <EGF1>
 F:144-175/Domain: EGF homology <EGF2>
 F:222-254/Domain: EGF homology <EGF2>
 F:261-292/Domain: EGF homology <EGF2>
 F:339-370/Domain: EGF homology <EGF3>
 F:416-449/Domain: EGF homology <EGF3>
 F:456-487/Domain: EGF homology <EGF4>
 F:494-525/Domain: EGF homology <EGF5>
 F:532-563/Domain: EGF homology <EGF6>
 F:607-638/Domain: EGF homology <EGF7>

A:Accession: S37238
A:Molecule type: mRNA
A:Residues: 1-502, 'L', 504-646 <HE2>
A:Cross-references: EMBL:Z22485, NID:9403306, PIDD:CA81259.1, PID:9403307

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1574 <NA>
A:Cross-references: EMBL:AB011532, NID:93449293, PIDN:BA32462.1, PID:93449293
A:Experimental source: strain Sprague-Dawley; brain
C:Genetics:
A:Gene: MEGF6

```

Db      142 VGGFYCRRCRPGYQIQDQDEKTCQDVDECAHNGGCGQHCRCVTPRGSYLCECKRGFRRLHNDGR 201
Qy      192 T---HSICT-----ACDES 202
          |      |  |  |

```

Db 202 TCLAISCTLNGSCQHCQVOLTIVTOHRKCCRPYOYLOEDGRRCVRSPCAEGNGGCMHI 261
QY 203 CKTCSGLTNRDCCGCEWGVY--DEGACVDYDECAAEPPCSAOFCKNANGSY----- 254
Db 262 CQELRLIAH--CG-CHPEYTOAADRKTEDEVCALGLAQ--AHGCLNTGSGFRCVCHA 316
QY 255 -----TCECDSSCV--GCTGEGPNCKECTISGAREHQ--CADV 291
Db 317 GYELGADGRCYRIEMELVNSCEANGSCSHGSHTSRPLCTCPRGLEDEDDKTCIDI 376
QY 292 DECLAETCYRKNECYNTPGSVVCYCPDGFEEEDAC 330
Db 377 DDC--ANSPPCQ--QACANTPGYECSCFAGYRLNTDGC 411

RESULT 38
A56136
Jagged protein precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change 11-Jan-2000
C:Accession: A56136
R:Lindseell, C.E.; Shawber, C.J.; Boulter, J.; Weimaster, G.
Cell 80, 909-917, 1995
A:Title: Jagged: a mammalian ligand that activates Notch1.
A:Reference number: A56136; MUID:95211842
A:Accession: A56136
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1220 <LIN>
A:Cross-references: GB:L38483
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:379-410/Domain: EGF homology <EGF1>
F:492-523/Domain: EGF homology <EGF2>
F:634-665/Domain: EGF homology <EGF2>

Query Match 11.2%; Score 224; DB 2; Length 1220;
Best Local Similarity 22.8%; Pred. No. 5.5e-07;
Matches 97; Conservative 42; Mismatches 101; Indels 186; Gaps 28;

QY 42 NOGAVD-----TAKKNGGNTAMEEKTLSKESSERLLEIGLCESS--DFECNQ 93
Db 158 HSGMINSRQWQTLKQNG-----IAHFE-YQIRV-----TCDDHYGFGCKKF 200
QY 94 LEAGEHLEAMWLKSEYPLDF--EMPCVTKLYCCSPGTGPPC--LACGGSGRPPSG 150
Db 201 CRPND-----DFGHYACDQNGKNTCEGMGMPCKNAICRQ-----CSP 241
QY 151 NGHSGDGSROGDSRCRHMGYQGPLCTDCM--DGYFSSLNE----- 191
Db 242 K-----HGSCKLPDCCRQYQMGGLYCDKCLPHPCGVHGTCEPQCICETNMGSQLCDK 296
QY 192 -----THSIT-----TACDE-----SK-TC 206
Db 297 DLNYGTHQPLNKGTSNTGPRDYQCSGPRGSGPNCIEAHACLSDPCHNRGSKETS 356
QY 207 SGLNROGCEGVGV-----LDE-----GAC-----V 229
Db 357 SGF---EC-ECSPGWTGTCSTNIDDCSPNNCSHGCTQDLVNGFKVCYCPQWTKTCQL 412
QY 230 DVDECAEPPCSAOFCKNANGSYTC-----EECDSSCVGCTG--EGPGNCKE--- 276
Db 413 DANCEAK--PCVNARSKNLIASTYCDCLPGWNGQNDININDLGGCQNDASRDLYN 470
QY 277 -----CISGTAHEGQCADVDECSLAETCYRKNECYNTPGSVVCYCPDGF-----EET 326
Db 471 GYRCICPGYAGDHCE--RDIDEC--ASNPL--NGHCONEINRFQCLCPTGFSGNLCQLD 526
QY 327 EDACVP 332
Db 527 IDYCEP 532

RESULT 39
A57172
probable hormone receptor EMRI precursor - human
C:Species: Homo sapiens (man)
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 11-Jan-2000
C:Accession: A57172
R:Baud, V.; Chissee, S.L.; Vlegas-Pequignot, E.; Diriong, S.; N'Guyen, V.C.; Roe, B.A.
Genomics 26, 334-344, 1995
A:Title: EMRI, an unusual member in the family of hormone receptors with seven transmembrane domains
A:Reference number: A57172; MUID:95324926
A:Accession: A57172
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-886 <BAU>
A:Cross-references: GB:X81479; NID:g784993; PIDN:CAA57232.1; PID:g784994
C:Genetics:
A:Gene: GDB:EMRI
A:Cross-references: GDB:378349; OMIM:600493
A:Map position: 19p13.3-19p13.3
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: transmembrane protein
F:136-170/Domain: EGF homology <EGF>

Query Match 11.1%; Score 222.5; DB 2; Length 886;
Best Local Similarity 27.9%; Pred. No. 5.3e-07;
Matches 79; Conservative 20; Mismatches 95; Indels 89; Gaps 17;

QY 130 PGTYPDCLACQGSQRPSCNGHSGDGSROGDSRCRHMGYQGPLCTDCM---DGYFS 186
Db 74 PGVNRKIDIDEC--SQSPQCGPNSCK---NLGGRKCSCLDFGSPGNDWPGAPGNS 129
QY 187 SLRNETHSICACDESCCTCGSLTNRDGE-----CEGVWVDEGACVDYD 232
Db 130 -----CTDINE---C--LTSRVCPHSDCVNSGYSQCVGFISRNSTCEDVN 174
QY 223 ECAAEPPCSAOFCKNANGSYTC-----EECDSSC---V 264
Db 175 EC-ADPRACPRHACNNTVGNYSFCNPFESSGHLSCQGLKASCEIDICTEMCPINS 233
QY 265 GCTGEGPNC--CKECTISGAREHQ-----CADVDECSLAETCYRKNECYNTPG 313
Db 224 TCTNTP-PSGYFC--TCHPFAFSPSQLNFTDQGVRECRDIDECROPSTC--GPNISICTNALG 290
QY 314 SYVCYCPDGFEEEDACVPAEAETEGESPQT---LPSRDL 353
Db 291 SYSCGCIYGFH-----PNEGSGKDGNSFCQRYLFXCKEDV 326

RESULT 40
KXBOS
Plasma protein S precursor - bovine
N:Alternate names: vitamin K-dependent protein S
C:Species: Bos primigenius taurus (cattle)
C:Date: 08-Aug-1987 #sequence_revision 26-Jan-1996 #text_change 16-Jul-1999
C:Accession: A24759; A23888
R:Dahlback, B.; Lundvall, A.; Stenflo, J.
Proc. Natl. Acad. Sci. U.S.A. 83, 4199-4203, 1986
A:Title: Primary structure of bovine vitamin K-dependent protein S.
A:Reference number: A24759; MUID:86233400
A:Accession: A24759
A:Molecule type: mRNA
A:Residues: 1-675 <DNA>
A:Cross-references: GB:M13044; NID:g163697; PIDN:AAA30757.1; PID:g163698
A:Note: Parts of this sequence, including the amino end of the mature protein, were determined from the complementary DNA sequence of the mature protein.
J. Biol. Chem. 261, 5111-5115, 1986
A:Title: Localization of thrombin cleavage sites in the amino-terminal region of bovine plasma protein S.
A:Reference number: A23888; MUID:86168236
A:Accession: A23888
A:Molecule type: protein
A:Residues: 42-141 <DA2>
C:Complex: in plasma forms a complex with C4b binding protein

C:Function:
 A:Description: a cofactor for activated protein C (EC 3.4.21.69); thrombin cleavage dest
 C:Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat hom
 C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; blood coagulation; carbox
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-41/Domain: propeptide #status predicted <PRO>
 F:26-85/Domain: Gla domain homology <Gla>
 F:42-65/Product: plasma protein S #status experimental <MAT>
 F:121-154/Domain: EGF homology <EG1>
 F:161-199/Domain: EGF homology <EG2>
 F:205-241/Domain: EGF homology <EG3>
 F:247-282/Domain: EGF homology <EG4>
 F:315-666/Domain: sex hormone-binding globulin homology <SHB>
 F:325-478/Domain: laminin G repeat homology <LGR>
 F:47-48-55-57-60-61-66-67-70-73-77/Modified site: gamma-carboxyglutamic acid (Glu) #stat
 F:58-63-88-113-121-134-126-143-145-154-161-175-171-184-186-199-205-217-212-226-228-241,2
 F:93-94/Cleavage site: Arg-Ala (thrombin) #status experimental
 F:111-112/Cleavage site: Arg-Ser (thrombin) #status experimental
 F:136/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
 F:177-219,258/Modified site: erythro-beta-hydroxyasparagine (Asn) #status experimental
 F:288-567,449-475,638-665/Disulfide bonds: #status experimental
 F:499/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:509/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.1%; Score 222; DB 1; Length 675;
 Best Local Similarity 26.4%; Pred. No. 4,6e-07;
 Matches 83; Conservative 29; Mismatches 98; Indels 104; Gaps 20;

OY 57 NTAMEKTSKYSSRIRLEILEGICSSD-----FECNOMLEAOEHEAMWLOL----- 108
 DB 43 NTLLEETKKNLER-----ECIEELCNKEAREIEENN---PELEYEPKYLGLIGSF 92
 OY 109 -----KSEYDLEFEMFCVKTILKYCCSGPTGYPDCLACQGSORPCSGNCHCS-G 156
 DB 93 RAGLFRAALSTNAYPDLNS--CVNAISDQCN-----LPENNDGWTCK 135
 OY 157 DGSROGDGSCRCRMGYOGLC---TDGMDGYFSSLRNETHSICTACDESKTSGLTNR 212
 DB 136 DG--QATFTICICKSGWQGEKCESDINECKDPV-----NINGCSCQICENTPG--SY 182
 OY 213 DCGCEGVNWL--DEGACVDVDECAAEPPPCSAOFCRKNANGSYTEBEDSSCVGCTGB 270
 DB 183 HC--SCNGVFMVSNKRKDCXDVCEVLPKSTICGTA--VCKNIPGDFEC----- 226
 OY 271 PGNCKRCIGYARE--HGOCADYDECSLAETCVRKNNCYNTPGSYVCV--PDGFEET 326
 DB 227 -----ECARGYKKNPYKSKCDDVDEC--AENLCA---QLCVNTPGYSCTCDGKRGFKLA 276
 OY 327 ED-----ACVP 332
 DB 277 ODKSCCAVPVCLP 290

RESULT 41
 T20125
 Hypothetical protein C50H2.3a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20125
 R:McMurray, A.
 submitted to the EMBL Data Library, June 1996
 A:Reference number: Z19226
 A:Accession: T20125
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-838 <WII>
 A:Cross-References: EMBL:Z73971; PIDN:CA98251.1; GSPDB:GN00023; CESP:C50H2.3a
 A:Experimental source: clone C50H2
 C:Genetics:
 A:Gene: CESP:C50H2.3a
 A:Map position: 5
 A:Intons: 31/1; 98/1; 162/1; 199/1; 244/1; 279/1; 323/1; 354/2; 391/1; 408/1; 458/1; 48

Query Match 11.1%; Score 222; DB 2; Length 838;
 Best Local Similarity 27.2%; Pred. No. 5.5e-07;
 Matches 75; Conservative 25; Mismatches 92; Indels 84; Gaps 17;

OY 117 EMEFCVTLKVCSPGTYGPDCLACOG-----GSQRPC-----SG 150
 DB 51 QMEFMDKQVEE--KKEFYH-----GCMGTKNRFSKQCKRYKMPVAVPDLCLLDAD 105
 OY 151 NCHSGSDGS-----ROQDSCRCRMGYOGLPCTDC--MDGYFSSL-----RNFT 192
 DB 106 QGCGDERNGHMYFEPNDSGECEKFFY-----GCGNDNKFYSLHMCRCRYCGERLSP 159
 OY 193 HSICFACD---ESCKTCSGLTNRDCECEVNWIDE--GACVDVDECAEPPPCSAOFCR 248
 DB 160 QIACHDCLRTSTCKNSNF--NTYC--ECRSYKYNIGECIDIDEGRGKAVCDRANCV 217
 OY 249 NANGSYTCECDSSCVG---CTGEGPG---NKECT-----SGYAREHGO 287
 DB 218 NEIGSYKC--ECMASYRGDKHCTYVGLGRSSIDCKDSMHATCMNGVCQCKEGYEGDGFN 276
 OY 288 CADVDECSLAETCVRKNNCYNTPGSYVCPPDPF 323
 DB 277 CTDVNCECLRPEMC--NKNAECTINREGSFTCTLEGY 311

RESULT 42
 A38261
 masking protein precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 21-Jan-2000
 C:Accession: A38261
 R:Tsuji, T.; Okada, F.; Yamaguchi, K.; Nakamura, T.
 Proc. Natl. Acad. Sci. U.S.A. 87, 8835-8839, 1990
 A:Title: Molecular cloning of the large subunit of transforming growth factor
 A:Reference number: A38261; MUID:91062373
 A:Accession: A38261
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1712 <TSU>
 A:Cross-References: GB:M55431; NID:9207285; PIDN:AA442235.1; PID:9207286
 C:Superfamily: unassigned EGF-related proteins; EGF homology; LDL receptor ligand-bl
 F:911-947/Domain: EGF homology <EGF>

Query Match 11.1%; Score 222; DB 2; Length 1712;
 Best Local Similarity 31.1%; Pred. No. 9.6e-07;
 Matches 68; Conservative 19; Mismatches 62; Indels 70; Gaps 15;

OY 150 GNGHCSGDSROGDGSCRCRMGYOGLPCTDCMDGYFSSLRNETHSICTACDESKTSGSL 209
 DB 877 GAGHCINLPVRY--TCIYEGYK-----FS-----EQQRKCIDIE--CAQAHL 917
 OY 210 TNRDCECE-----VGVNL--DEGA--CYDVDECAEPPPCSAOFCRKNANGSYTC 256
 DB 918 CSQ--GRCENTEGSLFICLPAGFIASEBGSNCIDVDEC--LRDPVCDGH--CINTAGAFRC 973
 OY 257 EEDDS-----SCVGTGCEPGN--CKECSGYAREHGQA 289
 DB 974 EYCDSGYRMSRRKHCHDIDIECLTPSTCPREOCY---NSPGSTQCVPTCTEGFRGNNGGL 1029
 OY 290 DVDECSLAETCVRKNNCYNTPGSYVCVPPDGFETED 328
 DB 1030 DVDEC--LQPKVCH--NGSCTNIEGYSWCSCHNGYSPTPD 1065

RESULT 43
 A49175
 Notch B protein - mouse (fragment)
 N:Alternate names: Notch homolog
 C:Species: Mus musculus (house mouse)
 C:Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999

C:Accession: A49175; PH570; S32113
R:Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A:Title: Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of
A:Reference number: A49175; MUID:93178563
A:Accession: A49175
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1203 <LAR>
A:Cross-references: EMBL:X68279; NID:9287989; PIDN:CAA48340.1; PID:9287990
A:Experimental source: embryo
A:Note: Sequence extracted from NCBI backbone (NCBI:126158)
C:Comment: This protein has many EGF repeats and 11n-12/Notch repeats.
C:Comment: This protein is one of the neurogenic proteins controlling the decision between
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:143-174/Domain: EGF homology <EGF1>
F:482-513/Domain: EGF homology <EGF1>
F:560-591/Domain: EGF homology <EGF1>
F:674-705/Domain: EGF homology <EGF2>
F:712-743/Domain: EGF homology <EGF3>
F:836-867/Domain: EGF homology <EGF3>

Query Match 11.0%; Score 221; DB 2; Length 1203;
Best Local Similarity 22.4%; Pred. No. 8.4e-07;
Matches 91; Conservative 44; Mismatches 113; Indels 158; Gaps 24;

OY 28 PTPCH-RCRLVDFKNGMGWDYAKKNGGNTAMEETLSKYES---SEIRLEITLGL 82
DB 259 PDPCHHGCGQGDIDSY-----TICNPGYMGALICSDIDECYSPCLNDGRCDLVNG- 311
OY 83 CESSDFECNQMLEAEHELEAMWLQKSEYDLEFEMFCVTKLKCCSPGTGPPDC-LACQ 141
DB 312 -----YQCN-----CQGTSLNCEINFD 330
OY 142 GGSORPCSGNCHSGDGRQDSCRCMHGYOGPLC-----TDCMDGYFSSLRNFT 192
DB 331 DCASIPFCM-HGVCDVGINRY--SCVCSPTGQRNCNIDIECASNPCKRG--ATCINDV 384
OY 193 HSICTADE-----SCKT-----CS-GILNRQCEGEVWV-----LDGGA 227
DB 385 NGFRICCEGPHHPSQYVNECLSNPCIGHNCTGSLGYKC-LCDACWGVNCEVDNE 443
OY 228 C-----VDVDECAAPPCSAAPFCCKNANGSYT 255
DB 444 CLSNPCQNGCTCNLYNVRCTCKKGFKGYNCQVNIDECASN--PCLNMGTCFDVSYGT 501
OY 256 C-----EECDSSCVGCTG---EGPGNCKE-----CISGYAREHGQC-ADVDSCS 295
DB 502 CHCMLPYGKNCQVYLACSPPCENAVKKEAPNFESFSLCAPGMOGKRCYVDDEC- 560
OY 296 LAERTCVAKKNCYTPGSSYVCVCGDF-----EETEDACVPAPE 335
DB 561 -ISKPCM-NGVCHNTQGSYVCECPGFGMDCEEDINDCLANPCQ 604

RESULT 44

MMHUND

nidogen precursor - human
N:Alternate names: entactin
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Oct-2000
C:Accession: A33322; A32437; A61367
R:Nagayoshi, T.; Sanborn, D.; Hickok, N.J.; Olsen, D.R.; Fazio, M.J.; Chu, M.L.; Knowlton
DNA 8, 581-594, 1989
A:Title: Human nidogen: complete amino acid sequence and structural domains deduced from
A:Reference number: A33322; MUID:90091745
A:Accession: A33322
A:Molecule type: mRNA
A:Residues: 1-1247 <NAG>
A:Cross-references: EMBL:M30269
R:Olsen, D.R.; Nagayoshi, T.; Fazio, M.; Mattel, M.G.; Passage, E.; Well, D.; Timpl, R.;
Am. J. Hum. Genet. 44, 876-885, 1989

A:Title: Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to
A:Reference number: A32437; MUID:89270475

A:Accession: A32437
A:Molecule type: mRNA
A:Residues: 667-1247 <OLIS>
A:Cross-references: EMBL:M27445; NID:9602466; PIDN:AAA57261.1; PID:9602467
A:Note: the authors translated the codon AAG for residue 966 as Cys
R:Fazio, M.J.; O'Leary, J.; Kaehner, V.M.; Chen, Y.Q.; Salita, B.; Uitto, J.
J. Invest. Dermatol. 97, 281-285, 1991
A:Title: Human nidogen gene: structural and functional characterization of the 5'-fla
A:Reference number: A61367; MUID:91302882

A:Accession: A61367
A:Molecule type: DNA
A:Residues: 1-28 <FAZ>
C:Comment: This protein is a basement membrane glycoprotein that forms a complex with
C:Genetics:
A:Gene: GDB:NID
A:Cross-references: GDB:120236; OMIM:131390
A:Map position: 1q43-1q43
C:Superfamily: nidogen; EGF homology; LDL receptor WYTD-containing repeat homology; t
C:Keywords: basement membrane; beta-hydroxyasparagine; calcium binding; cell binding;
protein

F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1247/Product: nidogen #status predicted <MAT>
F:390-425/Domain: EGF homology <EG1>
F:672-708/Domain: EGF homology <EG2>
F:702-704/Region: cell attachment (R-G-D) motif
F:714-750/Domain: EGF homology <EG3>
F:762-800/Domain: EGF homology <EG4>
F:806-839/Domain: EGF homology <EG5>
F:849-919/Domain: thyroglobulin type I repeat homology <THY1>
F:990-1032/Domain: LDL receptor WYTD-containing repeat homology <WY1>
F:1033-1075/Domain: LDL receptor WYTD-containing repeat homology <WY2>
F:1076-1120/Domain: LDL receptor WYTD-containing repeat homology <WY3>
F:1121-1160/Domain: LDL receptor WYTD-containing repeat homology <WY4>
F:1161-1197/Domain: LDL receptor WYTD-containing repeat homology <WY5>
F:1212-1243/Domain: EGF homology <EG6>
F:1289-296/Binding site: sulfate (Tyr) (covalent) #status predicted
F:729-819/Modified site: epsilon-beta-hydroxyasparagine (Asn) #status predicted
F:755/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of laminin) #status pr
F:1137/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.0%; Score 221; DB 1; Length 1247;
Best Local Similarity 27.1%; Pred. No. 8.7e-07;
Matches 54; Conservative 28; Mismatches 75; Indels 42; Gaps 7;

OY 174 GPLCTDCMDGYFSSLRNETHSICITACDESKTSGILTRNDGCEGEVWLDGACVDVE 233
DB 658 GPV-----REGSPDALQNCYIGTHGCDINACRPFRPQFCEGIGFRGDRCTYDIDE 713
OY 234 CAEPFPCSAAPFCCKNANGSYTCEEC-----DSCVCGCTGESP-----GNCK-- 275
DB 714 CSEQPSVCGSHITICNNHGFTRC-ECVEGYQPSDGTCAVAVDQHPINVCETGLHNCIP 772
OY 276 -----ECISGYAREHGQADVDCEGLAEKTCVKKKNCYTPGSSYVCVCP 320
DB 773 QRAQCITYGSSYTCSCLPGRFSGDQACADVDCE---QPSRCHPAFCYTNPGSTTCCK 829

OY 321 DGEETEDACVPAPEAAT 339

DB 830 PXYQGDGFRVCV-PGEVEKT 847

RESULT 45

A40043

notch protein homolog TAN-1 precursor - human
C:Species: Homo sapiens (man)
C:Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 13-Aug-1999
C:Accession: A40043
R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Skl
Cell 66, 649-661, 1991
A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromoso

A:Reference number: A40043; MUID:91347367
 A:Accession: A40043
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-2555 <ELL>
 A:Cross-references: GB:M73980
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:261-292/Domain: EGF homology <EGF1>
 F:494-525/Domain: EGF homology <EGF1>
 F:987-1018/Domain: EGF homology <EGF2>
 F:1149-1180/Domain: EGF homology <EGF3>
 F:1187-1218/Domain: EGF homology <EGF3>
 F:1233-1264/Domain: EGF homology <EGF3>
 F:1927-1959/Domain: ankyrin repeat homology <AN1>
 F:1960-1992/Domain: ankyrin repeat homology <AN2>
 F:1994-2026/Domain: ankyrin repeat homology <AN3>
 F:2027-2059/Domain: ankyrin repeat homology <AN4>
 F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 11.0%; Score 221; DB 2; Length 2555;
 Best Local Similarity 22.4%; Pred. No. 1.5e-06;

Matches 79; Conservative 20; Mismatches 105; Indels 148; Gaps 16;

QY 125 KVCSPFQYGPDCGLACQGSQRPSCSGNGHSCGDSRQDGSRCRHMGYQPLCTD----- 179
 DB 126 KCRCPGWSKSCQADPCASNPNCANGQCL---PFEASTYICHPSPFHGPTCRQDVNEC 182
 QY 180 -----CMGTFSSLNETHS---ICTA-----CDESCSTCS----- 207
 DB 183 GQKPRLCRHH--GTCHEVGSYRCVCRATHHTGPNCEPRYPVSPSPQNGTCRPTGDTV 240
 QY 208 -----GLTRDQGE-----CEVGWVLDGACV-----D 230
 DB 241 HECACLPGFTGQCEINIDDCPGNNCKNG-----GACVDSVNTYNGRCPEWTGQCTED 295
 QY 231 VDECAAEPPPCSAOFCCKNANGSYTC-----EBCDSCVCGCTGEG---PENCKE--- 276
 DB 296 VDECAAEPPPCSAOFCCKNANGSYTC-----EBCDSCVCGCTGEG---PENCKE--- 276
 QY 277 -----CISGVAREHQGC-----ADVDE 293
 DB 356 SFCECPHGTGLCHLNDACISNPNCEGSCNCTNPYNGKALICTCSGTYGPAACSDVD 415
 QY 294 CSIAERTCYARKNENCTNPSPGVCPCDGF-----EETEDACVP-PAEAET 339
 DB 416 CSIGANPCEHAGK-CINTLGSEFCQCLQGYTGPRCEIDVNECVSNPCQNDAT 466

RESULT 46
 S18188

notch protein homolog - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999

C:Accession: S18188

R:Weinmaster, G.; Roberts, V.J.; Lemke, G.

Development 113, 199-205, 1991

A:Title: A homolog of Drosophila Notch expressed during mammalian development.

A:Reference number: S18188; MUID:92111383

A:Accession: S18188

A:Molecule type: mRNA

A:Residues: 1-2531 <WEI>

A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

F:987-1018/Domain: EGF homology <EGF1>

F:1025-1056/Domain: EGF homology <EGF2>

F:1233-1264/Domain: EGF homology <EGF2>

F:1917-1949/Domain: ankyrin repeat homology <AN1>

F:1950-1982/Domain: ankyrin repeat homology <AN2>

F:1984-2016/Domain: ankyrin repeat homology <AN3>

F:2017-2049/Domain: ankyrin repeat homology <AN4>

F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 11.0%; Score 220; DB 2; Length 2531;
 Best Local Similarity 23.6%; Pred. No. 1.7e-06;
 Matches 76; Conservative 23; Mismatches 97; Indels 126; Gaps 17;

QY 126 VC-CSFGTYGPDCGLACQGSQRPSCSGNGHSCGDSRQDGSRCRHMGYQPLC- 177
 DB 163 ICQCPGPFHGPPTCRQDVNEC---SQNPGLCRHGGCTCHNE---IGSYRCACRATHHTGPHCE 216
 QY 178 -----TDCKMDGYFSSLNETHSICTA-----CDESCSTCSGLTRNDQGECEVM 221
 DB 217 LRPVPCSPSCQNGGTCRPTGDTTHCACLPFAGNCEENVDGCG-----NMCKNG- 269
 QY 222 VDEGACV-----DVDECAAEPPPCSAOFCCKNANGSYTC----- 256
 DB 270 -----GACVDSVNTYNGRCPEWTGQCTEDVDECAAEPPPCSAOFCCKNANGSYTC----- 325
 QY 257 -----EBCDSCVCGCTGEG---PENCKE-----CISGVAREHG 286
 DB 326 GWTGEDCSNIDDCASACFOGATCHDRVASFCECPHGTGLCHLNDACISNPNCEGS 385
 QY 287 QC-----ADVDECSIAERTCYARKNENCTNPSPGVCPCDGF----- 323
 DB 386 NCDTNPYNGKALICTCPRGTYGPAACSDVDVDECAAEPPPCSAOFCCKNANGSYTC----- 444
 QY 324 -----EETEDACVP-PAEAET 339
 DB 445 TGPRCEIDVNECVSNPCQNDAT 466

RESULT 47
 T08618

intrinsic factor-B12 receptor CUBILIN precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: T08618

R:Moestrup, S.K.; Kozyrak, R.; Kristiansen, M.; Kayser, J.H.; Rasmussen, H.H.; Braul

J. Biol. Chem. 273, 5235-5242, 1998

A:Title: The intrinsic factor-B12 receptor and target of teratogenic antibody

A:Reference number: T08618

A:Accession: T08618

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3623 <MO>

A:Cross-references: EMBL:AF022247; NID:g3834379; PIDN:AACT1661.1; PID:g3834380

C:Genetics:

A:Gene: CUBILIN

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra

F:1-20/Domain: signal sequence [status predicted <Sig>

F:21-3623/Product: intrinsic factor-B12 receptor CUBILIN #status predicted <Mat>

F:133-164/Domain: EGF homology <EGF1>

F:436-467/Domain: EGF homology <EGF>

Query Match 11.0%; Score 220; DB 2; Length 3623;
 Best Local Similarity 26.0%; Pred. No. 2.3e-06;
 Matches 64; Conservative 22; Mismatches 74; Indels 86; Gaps 12;

QY 120 CVKTL---KVCSPFQYGPDCGLACQGSQRPSCSGNGHSCGDSRQDGSRCRHMGY 172
 DB 187 CVNTVGSFRCDDCTPRDYGQCAKYNDCEQGSQLOLK----- 223
 QY 173 QGFLCTDCMDGYFSSLNETHSICTACDSCCKKCSGLTRNDQGECEVWVLDG--ACV- 229
 DB 224 -----HGI---CEDLQRYNHQPRPHC-ICDAGTTPRNNGISCTE 259
 QY 230 DVDECAAEPPPCSAOFCCKNANGSYTCSECDSSCVGCTGEGNCKECTSGVAREHGQCA 289
 DB 260 DKRCEGLQSPCEHQAQCNFTQGSFYCGACPKGMQ-----NGY-----EQO 301
 QY 290 DVDECSIAERTCYARKN-ENCTNPSPGVC-PCPDGEETEDACVP-----PAEAET 338

Db 302 DINECEINNGSGQAPLPLCLTPGSPSCNCPAGSGDGRVCTPVDICSIHNGCHPEA 361
Qy 339 TEGESP 344
Db 362 TCSSSP 367

RESULT 48

A:Gene: cell-fate determining gene Notch2 protein - rat
A:Accession: A49128
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A49128
R:Meinaster, G.; Roberts, V.J.; Lemke, G.
Development 116, 931-941, 1992
A:Title: Notch2: a second mammalian Notch gene.
A:Reference number: A49128; MUID:93202015
A:Accession: A49128
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2471 <ME1>
A:Experimental source: Schwann cell
A:Note: Sequence extracted from NCBI backbone (NCBI:127811)
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:264-295/Domain: EGF homology <EGX1>
F:799-830/Domain: EGF homology <EGF1>
F:877-908/Domain: EGF homology <EGX2>
F:1029-1060/Domain: EGF homology <EGF>
F:1067-1098/Domain: EGF homology <EGF3>
F:1153-1184/Domain: EGF homology <EGF3>
F:1191-1222/Domain: EGF homology <EGF4>
F:1876-1908/Domain: ankyrin repeat homology <AN1>
F:1809-1941/Domain: ankyrin repeat homology <AN2>
F:1943-1975/Domain: ankyrin repeat homology <AN3>
F:1876-2008/Domain: ankyrin repeat homology <AN4>
F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 10.9%; Score 218; DB 2; Length 2471;
Best Local Similarity 22.6%; Pred. No. 2.3e-06;

Matches 92; Conservative 43; Mismatches 112; Indels 160; Gaps 24;

Qy 28 PTPCH--RCRGLVDFKFNQGNVDTAKKNGGNTAMEEKLTKYS--SEIRLEILEGL 82
Db 576 PDPCHHGCGGIDSY-----TCICNGYMGALICSDQIDICSPCLNDRCLDVLNG- 628
Qy 83 CESSDFECNOMLEAOEHLLEAMWLQKSEYDLEFEMFCVTKLVCCSPGYGPDCLACQ 141
Db 629 --YOCN-----COPGTGLNCEINFD 647
Qy 142 GGSQRPCSGNGHSGDGRGDSRCRMGYQGPLCTDCMDGYFS--SLRNEHHS 194
Db 648 DCASNPL--HGACVDGIRY--SCVCSPTGTGRCNIDIDECASNPCRKATCLINDVNG 703
Qy 195 ICTACDE--SCRT-----CS-GLTNRDCECEVGVY--LDEGAC- 228
Db 704 FRCHCPREPHRSCYSQVNECLSPICNGCTGSLGSKC-ICDAGWVGINCVDKNCL 762
Qy 229 -----VDVDECAEPPPCSAOFCNNANGSYTC- 256
Db 763 SNPONGTGNLVNGYRCTCKKGFKGYNQVINDECASN--PCINOGTCLDVSAGYTC 820
Qy 257 -----EECDSCVCGTG--EGPGNCKE-----CISGARHSGQ--CADVDEC 294
Db 821 CMLEPTGKNCOTVLAPCSPPNCENANVCKEAPNFESFTCLCAPGV--OGRCYVDVDEC 877
Qy 295 SLAKTCVRKNENYNTPGSYVCVPDGF-----EETEDACVPPAE 335
Db 878 --VSKPCM--NNGICHTNNGSYMCBPPGFSGMDEBEDINDCLANCO 921

RESULT 49

T16860

hypothetical protein T13C2.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Aug-2000
C:Accession: T16860

R:Du, Z.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T13C2.
A:Reference number: 218591
A:Accession: T16860

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1357 <D02>

A:Cross-references: EMBL:U00030; NID:g1055164; PID:g1055166; PIDN:AAA81134.1; CESP:T1
A:Gene: CESP:T13C2.4
A:Intons: 38/2; 66/2; 115/2; 158/3; 233/1; 287/3; 325/3; 350/2; 432/2; 502/2; 525/1;
C:Superfamily: LDL receptor ligand-binding repeat homology
F:526-561/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:565-609/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:614-653/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:660-699/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:706-741/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:748-784/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:791-836/Domain: LDL receptor ligand-binding repeat homology <LDL7>

Query Match 10.8%; Score 217.5; DB 2; Length 1357;
Best Local Similarity 22.7%; Pred. No. 1.5e-06;

Matches 121; Conservative 43; Mismatches 141; Indels 229; Gaps 32;

Qy 1 MRLRRRAA--LGLLP-----LTLPLPAEAAKPP-----PCHR--CRGLVD 39
Db 441 VNLNRKSHGFLPEIYVTEERKLCLLPPRRKRSVYTONNTPIHKSVCQQLVYL 500
Qy 40 KFNQGNVDTAKKNGGNTAMEEKLTKYSSEIRLEILEGLCESSDFECNOMLEAO-- 97
Db 501 SSRTVREHVRVYFRLKNI---KTLRL--ECDATNSFCQDGRICIPMSWRCDIDICQNE 555
Qy 98 -----EHLLEAMWLQKSEYDLEFEMFCVTKLVCC-- 127
Db 556 EDEKNCPKVCAGEHKCG--EVKSARSLSLEFICIPKRWCVDEDFEDCKSDPECKNV 612
Qy 128 -----CS--PGTYG--PDCLACOGGSGORPCSGN--HSGDGRGDSRCRHM 170
Db 613 SCQKQKQPCGEBELSDYSLICETPTWVCG--QRDCT--NKKDQNKTSKSPDNNFQCSN 669
Qy 171 G--YQGPLC--TDCMDGYFSSLR--NET----- 192
Db 670 GNCIFKMWVCGDEBDCSDGDELTLAPSNCRVYVQCPGEMMKGSGECIPSRMCDAE 729
Qy 193 -----HS--ICTACDECK-----TCS-GLTNRDCE-- 215
Db 730 VDKDHSDEKNCVTAIQRCKLAEEFPACKASHNCINKAFVCDGELDCSDGSDDEDCADVRT 789
Qy 216 ECEVGV-----VWLDEGACVDVD----- 232
Db 790 ECKSGERTCPASYAYGAESEHNVICIPASSMNCNGEPPDGDDEKCMHTAPYQCKTE 849
Qy 233 -ECAAP-----PCSAQF--CKNANGSYTC-----EECDSCVCGCT-----GEG 270
Db 850 YECSTPLQCIEMSKLASAOFDCGDGMS--VCSQKTIEMCKPSSSEGCVRPSPFRGNN 908
Qy 271 PGNCKECSIGYAREHGOCADVDECSLAKTCVRKNENYNTPGSYVCVPDGF 324
Db 909 VCHCKD--GYKLENGQCIDINECEIA-GVC--DQICLNPISGRACCHAGYQ 955

RESULT 50

T23433

hypothetical protein K08C7.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T23433

R:Berks, M.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19740
A:Accession: T23433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3672 <WIL>
A:Cross-references: EMBL:Z70286; PIDN:CAA94293.1; GSPDB:GN00022; CESP:K08C7.3
A:Experimental source: clone K08C7
C:Genetics:
A:Gene: CESP:K08C7.3
A:Map position: 4
A:introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3; 3
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like R

Query Match 10.8%; Score 217.5; DB 2; Length 3672;
Best Local Similarity 23.1%; Pred. No. 3.4e-06;
Matches 93; Conservative 31; Mismatches 170; Indels 109; Gaps 20;
QY 31 CHR-CRGLVDKFNQGNVDTAKKN-----FGGNTAW--EETLSK-----YESS 71
DB 331 CERCCPGEFVK-QMGAATRAHNNFTCEACNCFRNSNCEYDAVDLNKQSIDSQNYEGG 388
QY 72 EIRLLEILBGLCE-----SSDFECNOMLEAQEHLBAMW-----LQKSEYPDLFEMFCV 121
DB 389 -----GVCKNCRENTEGVNCKNCSFGYFRPEGVYTNNEPQPKVCDCCDPDKHTGACA 439
QY 122 KTIKVC-CSPGTGPPCLACOG-----SORPSSNGHCS-GDGSROGDSGRCHMGYOG 174
DB 440 EETGKCECLPRFVGEDCDQASGYDAPKCKPCECNVNGTIGDCLPEDGQCPKAGFGG 499
QY 175 PLCTDCMDGYFFSSLRNETHSICTACDESCKTCGSLTNR-----DCGCEVGMVLD 224
DB 500 TRCETCADGTNTTACGVEVCDAATGSEHNGCSASTGQCECKPAYAGLSGDKQVGYFGD 559
QY 225 EGACVDVDECAEPPECSSAAQ---FCKNANGSYTCECD-----SSCVGCTGEGPG--- 272
DB 560 DCFKCNCDPMGTGEGVCDQDTTGQCLCEKGFAGDKCDIAFYGPNCACACDGAGITS 619
QY 273 -----NCKECISGY-----AREHGOCADVLD-EC-----S 295
DB 620 PECDATSGGCCPCNGNFTGRTCDKCAAGFYNYPDRCGCECLLSGAKGQTCDSNGQCYCKGN 679
QY 296 LAEKTQVRKNENCYNTPGSYVCV-PDGEFETEDAC--VPPAE 335
DB 680 FEGERCDRCRPNFYNFPICEECNCNPSGYTRDQGGCDKVSPE 722

Search completed: September 10, 2002, 11:11:54
Job time: 201 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 10:58:58 ; Search time 13.17 Seconds
(without alignments)
654.688 Million cell updates/sec

Title: US-09-905-075-2

Perfect score: 2005

Sequence: 1 MRLPRRAAGLPLLLLP.....AEAEATGEGSPQLPSREDL 353

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*

3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*

4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*

5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*

6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235.5	12.7	1253	3	US-08-479-722B-4
2	234.5	12.7	652	2	US-08-751-305-2
3	250	12.5	1833	3	US-08-479-722B-2
4	250	12.5	1833	5	PCT-US95-02251-18
5	246.5	12.3	1251	5	PCT-US95-02251-3
6	246.5	12.3	1252	1	US-08-199-780-3
7	246.5	12.3	1252	1	US-08-316-650-3
8	240.5	12.0	1394	6	517197-30
9	239	11.9	676	1	US-08-282-141-4
10	239	11.9	676	1	US-08-435-434-3
11	239	11.9	676	1	US-08-435-436-3
12	239	11.9	676	2	US-08-438-863-3
13	239	11.9	676	2	US-08-438-864-3
14	239	11.9	676	4	US-08-438-862-3
15	239	11.9	676	4	US-08-628-747-3
16	239	11.9	676	4	US-08-402-253-3
17	239	11.9	676	4	US-08-443-866B-3
18	236	11.8	799	2	US-08-525-940-23
19	236	11.8	799	2	US-08-976-838-23
20	236	11.8	881	2	US-08-525-940-21
21	236	11.8	881	2	US-08-976-838-21
22	236	11.8	881	2	US-08-525-940-18
23	236	11.8	915	2	US-08-976-838-18
24	230.5	11.5	288	2	US-08-525-940-15
25	230.5	11.5	288	2	US-08-976-838-15
26	227.5	11.3	635	1	US-07-907-190-1
27	227.5	11.3	635	1	US-07-985-691-2

ALIGNMENTS

28	227.5	11.3	635	1	US-08-436-804-2	Sequence 2, Appl1
29	227.5	11.3	635	1	US-08-267-387-2	Sequence 2, Appl1
30	224	11.2	652	6	5258288-4	Patent No. 5258288
31	224	11.2	1219	4	US-08-882-046-5	Sequence 5, Appl1
32	223.5	11.1	288	1	US-08-368-852-15	Sequence 18, Appl1
33	223	11.1	2523	1	US-08-185-432-18	Sequence 3, Appl1
34	222.5	11.1	886	3	US-09-110-116-3	Sequence 17, Appl1
35	222	11.1	2556	3	US-08-185-432-17	Sequence 20, Appl1
36	222	11.1	2556	3	US-08-083-590A-20	Sequence 20, Appl1
37	222	11.1	2556	3	US-08-532-384-20	Sequence 20, Appl1
38	220.5	11.0	1010	4	US-08-882-046-7	Sequence 7, Appl1
39	220.5	11.0	1036	4	US-09-068-740A-7	Sequence 6, Appl1
40	220.5	11.0	1187	4	US-09-068-740A-7	Sequence 7, Appl1
41	220.5	11.0	1218	2	US-08-400-159-6	Sequence 6, Appl1
42	220.5	11.0	1218	3	US-08-611-729A-6	Sequence 6, Appl1
43	220.5	11.0	1218	4	US-08-882-046-2	Sequence 2, Appl1
44	220.5	11.0	1218	4	US-09-214-278-7	Sequence 7, Appl1
45	220.5	11.0	1218	4	US-09-068-740A-11	Sequence 11, Appl1
46	218.5	10.9	675	6	5258288-1	Patent No. 5258288
47	216.5	10.8	2703	1	US-08-185-432-19	Sequence 19, Appl1
48	215.5	10.7	2471	1	US-08-185-432-16	Sequence 16, Appl1
49	215.5	10.7	2471	1	US-08-083-590A-19	Sequence 19, Appl1
50	215.5	10.7	2471	3	US-08-532-384-19	Sequence 19, Appl1
51	213	10.6	673	1	US-08-282-141-3	Sequence 3, Appl1
52	213	10.6	673	1	US-08-435-434-1	Sequence 1, Appl1
53	213	10.6	673	1	US-08-435-436-1	Sequence 1, Appl1
54	213	10.6	673	2	US-08-438-863-1	Sequence 1, Appl1
55	213	10.6	673	2	US-08-438-864-1	Sequence 1, Appl1
56	213	10.6	673	2	US-08-438-862-1	Sequence 1, Appl1
57	213	10.6	673	4	US-08-628-747-1	Sequence 1, Appl1
58	213	10.6	673	4	US-08-402-253-1	Sequence 1, Appl1
59	213	10.6	673	4	US-08-443-866B-1	Sequence 1, Appl1
60	211.5	10.5	810	2	US-08-820-170A-34	Sequence 34, Appl1
61	211.5	10.5	810	2	US-09-055-699-34	Sequence 34, Appl1
62	211.5	10.5	810	4	US-09-273-565-34	Sequence 34, Appl1
63	211.5	10.5	810	4	US-09-565-538-34	Sequence 34, Appl1
64	209.5	10.4	448	2	US-08-884-072-1	Sequence 1, Appl1
65	209.5	10.4	448	4	US-09-212-168-1	Sequence 1, Appl1

RESULT 1
US-08-479-722B-4
Patent No. 6074840
GENERAL INFORMATION:
APPLICANT: Bonadio, Jeffrey
APPLICANT: Yin, Wushan
TITLE OF INVENTION: LATENT TGF (BINDING PROTEIN (LTBP)
TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Williams, Morgan & Amerson
STREET: 7676 Hillmont, Suite 250
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77040
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479, 722B
FILING DATE: 07-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US95/02251
FILING DATE: 21-FEB-1995
PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: 4100.000500/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 934-7000
; TELEFAX: (713) 934-7011
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-479-722B-4

```

```

Query Match          12.7%; Score 255.5; DB 3; Length 1253;
Best Local Similarity 24.0%; Pred. No. 3e-12;
Matches 102; Conservative 40; Mismatches 108; Indels 175; Gaps 25;

```

```

QY 20 PABEAAKRPPTCHRCRGVDFKNOGVDTAKKNFGGNTAMEKTLKYSSEIRLEIL 79
DB 518 PYELISRPSP-----PTFHRFLPDLPP-----SRSAVEIAPQVETDECRINQNI 564
QY 80 --EGLC--ESSDF--ECNQMLEAQBHEHLEAMWQLKSEYDLEFEMCYVKLAKCCSPGTY 133
DB 565 CGHGCVPPGSPSDSCHMNGYRSHPOH-----YCVDVNE--CEAEPG 605
QY 134 GPDLCAQ--GGSORPCSGNGHS-----GDGSRQ-----GDG----- 164
DB 606 GPKGKICMNTGSSY-----NCHGNRGYRLHYGAGRSQVDLNCARPHLCGGDGFCINFP 660
QY 165 ---SCHMNGY-----QGPLCTD-----CMDGYFSSLNETHSI--CTACDESCKTC 206
DB 661 GHYKCMCYPGYRLKASRPICEDIDECRDPSTCPDG--KCEMKPSFKICACQPGYRSQ 717
QY 207 SGLTNDCCGECEYGVLDDEACVDECAAEPPPCSAQCFKANKANGSYNC----- 256
DB 718 GG-----GACRPVNEC--SEGTPCSPG--WCENLPDSYRCTCAQGIPTRT 758
QY 257 -----EECDSSCY--GCTGEGGNGCK--ECISGY--AREHGQCADVDECSLAEK--- 299
DB 759 GRLSCIDVDECEAGKYCQDQICITNTPGSPQCCCLSGYHLSDRSRCEIDIDECFPAACTG 818
QY 300 -TCVRNEN-----CYNTPGSSVVCVCPDGF 323
DB 819 GDCINTNGSYRCLPLGHLRVGGRCKKDIDECSDPGLCLPHACENLQGSYVCVDEGF 878
QY 324 EETED 328
DB 879 TLTQD 883

```

```

RESULT 2
US-08-751-305-2
; Sequence 2, Application US/08751305
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Tennet et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Metherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-751-305-2

```

```

Query Match          12.7%; Score 254.5; DB 2; Length 652;
Best Local Similarity 30.7%; Pred. No. 1.8e-12;
Matches 78; Conservative 27; Mismatches 88; Indels 61; Gaps 15;

```

```

QY 95 EAOEHLLEAMWQLKSEYDLEFEMCYVKLAKCCSPGTYGPDCLACQGSORPCS-GNGH 153
DB 227 ESKDKDTQSHYFLCKEKADVPFW-----GSSGPLCVSPKYG---CNFNNG 270
QY 154 CSGDGSROGDS--CRCHMGYQGPLCTCDMDGYFSSLNETHSICVACDESCKTCGLTN 211
DB 271 CHDDCFEGGDSGLCCRGGRFR--LDDLVY---CASRNPCSSPPRGATCYLGHGKN 325
QY 212 RDCGECEYGVLDDEGA--CYVDDECAAEPPPCSAQCFKANKANGSYNC----- 324
DB 326 YTC--RCPOGYQLDSSQLDCVDEDEC--QDSPC--AQECVTPPGFRC-----CWGYEP 375
QY 269 EGRGNKECTSGYAREHGQCADVDECSLAEKTCVRNENKYNTPGSSVVCVCPDGF----- 324
DB 376 GGG-----EGACQDVDECALGRSPCA---QGTNTDGSFHCSCERGVLAGE 420
QY 325 -----ETEDACVPP 333
DB 421 DGTQCDVDECVGP 434

```

```

RESULT 3
US-08-479-722B-2
; Sequence 2, Application US/08479722B
; Patent No. 6074840
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Yin, Wushan
; TITLE OF INVENTION: LATENT TGF ( BINDING PROTEIN (LTBP)
; TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Williams, Morgan & Amerson
; STREET: 7676 Hillmont, Suite 250
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/479,722B
FILING DATE: 07-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US95/02251
FILING DATE: 21-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,650
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fussey, Shelley P.M.
REGISTRATION NUMBER: 39,458
REFERENCE/DOCKET NUMBER: 4100, 000500/FUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (713) 934-7000
TELEFAX: (713) 934-7011
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1833 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-479-722B-2

Query Match 12.5%; Score 250; DB 3; Length 1833;
Best Local Similarity 24.8%; Pred. No. 1.2e-11;
Matches 82; Conservative 22; Mismatches 94; Indels 132; Gaps 16;

QY 128 CSPG-----TYGPDCLACGGSGORPCSGNGHSGDSSROGDSRCRHMGY-----QG 174
DB 883 CSPGQLHPSDYCTDDNECM---RNPCEGRRCV---NSVGSYSCLCYPGYTLVTLGDT 936
QY 175 PLCTD-----CMDGYFSLRNETHSIC-----TACDESC 203
DB 937 QECODIDCEQPGVCSGRCSTESYHCECDRGYIMVRKHCQDINCRHHPGTCPPDRC 996
QY 204 KTCGSLTRNDCGCEGVNVLDEGACVDVE-----CAEP----- 238
DB 997 VNSPG--SYTCLACEGEGYVGSGSCVDVNECLTPGICTHGRGINNEGSFRCSCPEGYEV 1054
QY 239 -----PCSA-----AOFCKNANGSYTCECDSS-----SCVG----- 265
DB 1055 PDKKGRDVEDCASRASCPGTGLCLNTESFTCSACOSGYWNEDGTACEDDDECAFPQVC 1114
QY 266 ---CTGE-GPNCNKECISGYARE--HGQCADVDECSLAETCVARNKNCYNTPGSYVCV 318
DB 1115 PTGVTNTVGSFSCKDCQOGRPNPLGNRCEDVDECEGPQSSC--RGGECKNTESYQCL 1172
QY 319 CPDGEFT-----EDACVPPAE 335
DB 1173 CHQGFQLVNGTMCEDVNECVGEHCAPHGE 1202

RESULT 4
PCT-US95-02251-18
Sequence 18, Application PC/TUS9502251
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
NUMBER OF INVENTION: CELLS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: PatentIn Release #1.0, Version
SOFTWARE: #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02251
FILING DATE: CONCURRENTLY HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC009P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ. ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1833 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02251-18

Query Match 12.5%; Score 250; DB 5; Length 1833;
Best Local Similarity 24.8%; Pred. No. 1.2e-11;
Matches 82; Conservative 22; Mismatches 94; Indels 132; Gaps 16;

QY 128 CSPG-----TYGPDCLACGGSGORPCSGNGHSGDSSROGDSRCRHMGY-----QG 174
DB 883 CSPGQLHPSDYCTDDNECM---RNPCEGRRCV---NSVGSYSCLCYPGYTLVTLGDT 936
QY 175 PLCTD-----CMDGYFSLRNETHSIC-----TACDESC 203
DB 937 QECODIDCEQPGVCSGRCSTESYHCECDRGYIMVRKHCQDINCRHHPGTCPPDRC 996
QY 204 KTCGSLTRNDCGCEGVNVLDEGACVDVE-----CAEP----- 238
DB 997 VNSPG--SYTCLACEGEGYVGSGSCVDVNECLTPGICTHGRGINNEGSFRCSCPEGYEV 1054
QY 239 -----PCSA-----AOFCKNANGSYTCECDSS-----SCVG----- 265
DB 1055 PDKKGRDVEDCASRASCPGTGLCLNTESFTCSACOSGYWNEDGTACEDDDECAFPQVC 1114
QY 266 ---CTGE-GPNCNKECISGYARE--HGQCADVDECSLAETCVARNKNCYNTPGSYVCV 318
DB 1115 PTGVTNTVGSFSCKDCQOGRPNPLGNRCEDVDECEGPQSSC--RGGECKNTESYQCL 1172
QY 319 CPDGEFT-----EDACVPPAE 335
DB 1173 CHQGFQLVNGTMCEDVNECVGEHCAPHGE 1202

RESULT 5
PCT-US95-02251-3
Sequence 3, Application PC/TUS9502251
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
NUMBER OF INVENTION: CELLS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433

```

CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02251
FILING DATE: CONCURRENTLY HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC009P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02251-3

```

```

Query Match 12.3%; Score 246.5; DB 5; Length 1251;
Best Local Similarity 23.5%; Pred. No. 1.5e-11;
Matches 100; Conservative 41; Mismatches 109; Indels 175; Gaps 25;

```

```

QY 20 PAPEAKKPTPCRCRGIVDKFNGMVDPAKKNFGGNTAMEKTLKYESSEIRLLEIL 79
DB 517 PYELISRPSP-----PTFRRLPDLPP-----SRSAVEIAPTQVETDECRLNOMI 563
QY 80 --EGIC--ESSDF--ECNOMLEAOEHLBAMWLQKSEYPLDFEFVCYTKLVCCSPGY 133
DB 564 CGHGQCVPEPSPDYSCHCNAGYRSHPOHR-----YCVDAVNE--CEAEPG 604
QY 134 GPCLACQ--GGSQPCSGNGHCS-----GDGSRQ-----GDG----- 164
DB 605 GPGKIGCMNTGGSY-----NCHCNNGRYRLHVAGGRSCVDLNECAKPHILCGDGGFCINFP 659
QY 165 ---SCRCHNGY---QGPLCTD-----CMDGYFSSLNETHSI--CTACDESCRTG 206
DB 660 GHKCMCYGYRLKASRPICEDIDECRDPSTCPDG--KCEKPKGSFKCIACQPGYRSQ 716
QY 207 SGLTRNDGCECEVGNVLDGACVYDECAEAPPSPCSAOFCKNANGSYTC----- 256
DB 717 GG-----GACRDVNEC--SEGTPCSPG--WCEKLPSTYCTCAOGIRTRT 757
QY 257 -----EEDSSCY--GCTGEGPGNCK--ECISGY--AREHGOCAVDDECSLAEK--- 299
DB 758 GRLSCIDVDDCGAGKVCODGICTNTPGSPQCCLSGTHLSRDRSCEDIDEDCPFAACIG 817
QY 300 -TCVRKNEN-----GACRDVNEC--SEGTPCSPG--WCEKLPSTYCTCAOGIRTRT 758
DB 818 GRLSCIDVDDCGAGKVCODGICTNTPGSPQCCLSGTHLSRDRSCEDIDEDCPFAACIG 817
QY 324 EETED 328
DB 878 TLTPD 882

```

```

RESULT 6
US-08-199-780-3
Sequence 3, Application US/08199780
Patent No. 5763416
GENERAL INFORMATION:
APPLICANT: Bonadio, Jeffrey
APPLICANT: Goldstein, Steven A.
TITLE OF INVENTION: Gene Transfer Into Bone Cells
TITLE OF INVENTION: And tissues
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1252 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-199-780-3

```

```

Query Match 12.3%; Score 246.5; DB 1; Length 1252;
Best Local Similarity 23.5%; Pred. No. 1.5e-11;
Matches 100; Conservative 41; Mismatches 109; Indels 175; Gaps 25;

```

```

QY 20 PAPEAKKPTPCRCRGIVDKFNGMVDPAKKNFGGNTAMEKTLKYESSEIRLLEIL 79
DB 518 PYELISRPSP-----PTFRRLPDLPP-----SRSAVEIAPTQVETDECRLNOMI 564
QY 80 --EGIC--ESSDF--ECNOMLEAOEHLBAMWLQKSEYPLDFEFVCYTKLVCCSPGY 133
DB 518 PYELISRPSP-----PTFRRLPDLPP-----SRSAVEIAPTQVETDECRLNOMI 564
QY 80 --EGIC--ESSDF--ECNOMLEAOEHLBAMWLQKSEYPLDFEFVCYTKLVCCSPGY 133
DB 565 CGHGQCVPEPSPDYSCHCNAGYRSHPOHR-----YCVDAVNE--CEAEPG 605
QY 134 GPCLACQ--GGSQPCSGNGHCS-----GDGSRQ-----GDG----- 164
DB 606 GHKCMCYGYRLKASRPICEDIDECRDPSTCPDG--KCEKPKGSFKCIACQPGYRSQ 716
QY 207 SGLTRNDGCECEVGNVLDGACVYDECAEAPPSPCSAOFCKNANGSYTC----- 256
DB 718 GG-----GACRDVNEC--SEGTPCSPG--WCEKLPSTYCTCAOGIRTRT 758
QY 257 -----EEDSSCY--GCTGEGPGNCK--ECISGY--AREHGOCAVDDECSLAEK--- 299
DB 759 GRLSCIDVDDCGAGKVCODGICTNTPGSPQCCLSGTHLSRDRSCEDIDEDCPFAACIG 818
QY 300 -TCVRKNEN-----GACRDVNEC--SEGTPCSPG--WCEKLPSTYCTCAOGIRTRT 758

```

Db 819 GDCINTNGSYRCLCPRLGRLVGRKCKKDIDECSDPDLCLPHACENLQGSYVCVCDGCF 878
QY 324 EETED 328
Db 879 TLTD 883

RESULT 7

US-08-316-650-3
; Sequence 3, Application US/08316650
; Patent No. 5942496
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Roesler, Blake J.
; APPLICANT: Goldstein, Steven A.
; APPLICANT: Lin, Mushan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS
; NUMBER OF INVENTIONS: FOR STIMULATING BONE CELLS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: US/08/316,650
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-316-650-3

Query Match 12.3% Score 246.5; DB 2; Length 1252;

Best Local Similarity 23.5% Pred. No. 1.5e-11;
Matches 100; Conservative 41; Mismatches 109; Indels 175; Gaps 25;

QY 20 PAPERAKPTCHRCRGVLDKFNQGMVDTAKKNFGGNTAMEEKLTSKESSEIRLLEIL 79
Db 518 PYPELISRSP-----PFHRLPLRP-----SRSAVIAIATQYTEIDECNLONI 564
QY 80 --EGIC--ESSDF--ECNOMLEAOEHLLEAMWLQKSEYDPLFEWFCVYKTLKVCSPGT 133
Db 565 CGHGQCVGSPDYSCHCNAGYRSHQHR-----YCVDAVE--CEAEP 605
QY 134 GPDLACQ--GSGRQPCSGNGHCS-----GDSGRQ-----GDS----- 164
Db 606 GPGKICNNTGGSY-----NCHCNNGYRLHVGAGGRCVDLNECAKPHLCGGGCGFCINFP 660
QY 165 ---SCRMHGY-----QGPLCTD-----CMDGYFSSLNETHSI--CTACDESCCTC 206

Db 661 GHYKNCVPGYRLKASRPICEDIDECRDPTSTCPDG---KCENRKGSFKCIACQPGVRSQ 717
QY 207 SGLTRNRCGEGCEVGVNIDEGACVYDDECAAEPPPCSAOFCNKANGSTTC----- 256
Db 718 GG-----GACRDVNEC--SGTTPCSPG--WCEKLPGSYRCAOIGIRTRT 758
QY 257 -----EECDSSCV---GCTGEGPNC--ECISGY---AREHGQCADVDECSLAEK--- 299
Db 759 GRLSCIDVDCEAGKVCODGICTNTPGSFQCCLSGHLHSDRSCEDIDECDFPAACIG 818
QY 300 -TVARKNEN-----CYNTPGSYVCVCPDGF 323
Db 819 GDCINTNGSYRCLCPRLGRLVGRKCKKDIDECSDPDLCLPHACENLQGSYVCVCDGCF 878
QY 324 EETED 328
Db 879 TLTD 883

RESULT 8

5177197-30
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO:30:
; LENGTH: 1394
; 5177197-30

Query Match 12.0% Score 240.5; DB 6; Length 1394;

Best Local Similarity 27.1% Pred. No. 5.2e-11;
Matches 85; Conservative 22; Mismatches 96; Indels 111; Gaps 19;

QY 67 KYESSEIRLLEILEGLCSSEDFECNOMLEAOEHLLEAMWLQKSEYDPLFEWFCVYKTLK 126
Db 655 EYCDSGYRMTQ--RGRCEDID--ECLNPSTCPDEQ-----CVN----- 688
QY 127 CCSPTTGPDLACQGGSGRQPCSGNGHC-----SGDGR--QDGSRCRMHGY 172
Db 689 --SPGSY--QCVPTCEGFR---GWNQCILDVDECLEPVMVCANGDCSNLEGSYMSCCHKGY 741
QY 173 QGPLCTDMDGYFSSLNETHSICTACDE-----SKTCSGLNRRGCEVEGVW 222
Db 742 -----TRPDHKNCHDIDECQGNLVCNGCKNTEGSRFTCTGQ---GYQ 783
QY 223 LD--EGACVDYDECAAEPPPCSAOFCNKANGSYTCECDSSCVCTEGPNCKECISG 280
Db 784 LSAKQDCEDIDEC--QHHLCAHQ--CRNTEGSPFC--VCDQ---GYRASGLD----- 830
QY 281 YAREHGQCADVDECSLAEKTCVRKNENCTNPGSYVCVCPDGF-----ETE 327
Db 831 -----HCEIDINECLEDRSVQR--GDCINTAGSYDCTCPDGFQDNNKTCODINECEHP 882
QY 328 DACVPAEAETEG 341
Db 883 GLCGPQGECLNTEG 896

RESULT 9

US-08-282-141-4
; Sequence 4, Application US/08282141
; Patent No. 5538861
; GENERAL INFORMATION:
; APPLICANT: Schneider, Claudio
; APPLICANT: Varnum, Brian

APPLICANT: Avanzi, Giancarlo
APPLICANT: Brancolini, Claudio
APPLICANT: Manfioletti, Guido
TITLE OF INVENTION: Stimulating Factor for the AXL Receptor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Amgen Inc.
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: United States
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/282,141
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-282-141-4

Query Match 11.9%; Score 239; DB 1; Length 676;
Best Local Similarity 25.8%; Pred. No. 3.1e-11;
Matches 92; Conservative 37; Mismatches 114; Indels 114; Gaps 22;
QY 12 LPLLLLPAPPAKPPCHRCGLVDFKNOGMDTAKNFGGNTAMEKTLSTKYS 71
DB 12 LACLLLVLPVSEA-----NLISKQASQVLYKRR---ANSLLEETKOGNLER 56
QY 72 EIRLELEIEGCESSDFECNOMLE--AOEHLBAMWL-----OLKSEYP 114
DB 57 -----ECIEELCNKE--EAREVFENDPEIDYFPYKYLCLNSFOTGLFTAROSTNAP 109
QY 115 LPEMFVCKTLKVCSPGTGYPDCLACQGSQPRCSGNHCS--GDGSRQDSCRCRMGY 173
DB 110 LRS--CVNAIPDQCP-----LPCNEDGYMSCKDG--KASFTCTCKPQM 150
QY 174 GPLC-----TDCMDGYFSSLRNETHSICITACDESKTCSGLTNRDCGECEVGNVL--DEGA 227
DB 151 GERCEFDINECKD-----PSNINGGCSQICDNTPG--SYHC--SKNGFVMLSNKK 198
QY 228 CAVNDECAAPPPCSAOPCKNANGSYTCEBDCSCVGTGEGPNCCKECTISGYA--REH 285
DB 199 CKVDDECSLKPSICGTA--VCKNIPDQEC-----ECPEGYRYNKK 238
QY 286 GGCADVDECSLAETKCVKRNENCYNTPGSYVCV--PDGEETED-----ACVP 332
DB 239 KSCCEDIDCS--ENMCA---QLCVNYPGGTYCYCDGKKGFKLAQDKSCCEVSVCLP 290

RESULT 10
US-08-435-434-3
Sequence 3, Application US/08435434
Patent No. 5714385
GENERAL INFORMATION:
APPLICANT: Mather, Jennie P.
APPLICANT: Li, Ronghao
APPLICANT: Chen, Jian
TITLE OF INVENTION: ISOLATING AND CULTURING SCHWANN CELLS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.
STREET: 460 Point San Bruno Blvd

CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,434
FILING DATE: 10-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 946-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-435-434-3

Query Match 11.9%; Score 239; DB 1; Length 676;
Best Local Similarity 25.7%; Pred. No. 3.1e-11;
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;
QY 11 LPLLLLPAPPAKPPCHRCGLVDFKNOGMDTAKNFGGNTAMEKTLSTKYS 70
DB 11 LACLLLVLPVSEA-----NLISKQASQVLYKRR---ANSLLEETKOGNLER 56
QY 71 SEIRLELEIEGCESSDFECNOMLE--AOEHLBAMWL-----OLKSEYP 113
DB 57 -----ECIEELCNKE--EAREVFENDPEIDYFPYKYLCLNSFOTGLFTAROSTNAP 108
QY 114 LPEMFVCKTLKVCSPGTGYPDCLACQGSQPRCSGNHCS--GDGSRQDSCRCRMGY 172
DB 109 DLRS--CVNAIPDQCP-----LPCNEDGYMSCKDG--KASFTCTCKPQM 149
QY 173 GPLC-----TDCMDGYFSSLRNETHSICITACDESKTCSGLTNRDCGECEVGNVL--DEG 226
DB 150 QGERCEFDINECKD-----PSNINGGCSQICDNTPG--SYHC--SKNGFVMLSNKK 197
QY 227 CAVNDECAAPPPCSAOPCKNANGSYTCEBDCSCVGTGEGPNCCKECTISGYA--RE 284
DB 198 DCKVDDECSLKPSICGTA--VCKNIPDQEC-----ECPEGYRYNKK 237
QY 285 HGCADVDECSLAETKCVKRNENCYNTPGSYVCV--PDGEETED-----ACVP 332
DB 238 SKCEDIDCS--ENMCA---QLCVNYPGGTYCYCDGKKGFKLAQDKSCCEVSVCLP 290

RESULT 11
US-08-435-436-3
Sequence 3, Application US/08435436
Patent No. 5721139
GENERAL INFORMATION:
APPLICANT: Mather, Jennie P.
APPLICANT: Li, Ronghao
APPLICANT: Chen, Jian
TITLE OF INVENTION: ISOLATING AND CULTURING SCHWANN CELLS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.

APPLICANT: Li, Ronghao
 TITLE OF INVENTION: RSE RECEPTOR ACTIVATION
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: patin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/438,864
 FILING DATE: 10-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/402253
 FILING DATE: 10-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Wendy M. Lee
 REGISTRATION NUMBER: 00,000
 REFERENCE/DOCKET NUMBER: 929P1-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1994
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 676 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-438-864-3

Query Match 11.9%; Score 239; DB 2; Length 676;
 Best Local Similarity 25.7%; Pred. No. 3.1e-11;
 Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;

```

QY 11 LPLLLLPAPPAKPPCHRCRGLVDKFNQGVDTAKKNGGNTAMEKTLSTKYES 70
DB 11 LLAALLLVLPVSEA-----NFLSKQASQVLVRRR--ANSLLLETKQGNLER 56
QY 71 SEIRLEILEIGCESSDPECNMLE--AOEHLLEAMWL-----QLKSEYP 113
DB 57 -----ECIEELCNKE--EAREVFENDPETDYFPRKYLIVCLRSFOTGLFTAROSTNAYP 108
QY 114 DLEFMCVTKLVKCSPGTYGPDCLACQGSQSPGSGNGHCS-GDGSROGDSRCRCHMGY 172
DB 109 DLRS--CVNAIPDQSP-----LPCNEDGYMSCKRG--KASTCTCKRGW 149
QY 173 QGPLC---TDCMDGYFSSLRNETHSICTACDESKTSGGLTNRDCGEGEYVWL--DEG 226
DB 150 QGEKCEFDINECKD-----PSNINGGCSQICDNTPG--SYHC-SCKNGFVLMISNKK 197
QY 227 ACVDVDECAEPPPCSAOFCNANGSYTCECDSSVCCTGEGNCKECLISGTA--RE 284
DB 198 DCKDVDECSLKPSIGTA--VCKNIPGDFEC-----ECPEGRYNLIK 237
QY 285 HQGCAADVDCSLAEKTCVKKNNENYNTPGSYVCV--PDGFETED-----ACVP 332
DB 238 SKSCEDIDCS--ENMCA---OLCVNYPGGTYCYCDGKKGKFLADQKSCYVSVCLP 290
  
```

RESULT 14
 US-08-438-862-3
 ; Sequence 3, Application US/08438862
 ; Patent No. 6033660
 ; GENERAL INFORMATION:
 ; APPLICANT: Mather, Jennie P.

APPLICANT: Li, Ronghao
 TITLE OF INVENTION: ISOLATING AND CULTURING SCHWANN CELLS
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: patin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/438,862
 FILING DATE: 10-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Lee, Wendy M.
 REGISTRATION NUMBER: 00,000
 REFERENCE/DOCKET NUMBER: 946-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1994
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 676 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-438-862-3

Query Match 11.9%; Score 239; DB 3; Length 676;
 Best Local Similarity 25.7%; Pred. No. 3.1e-11;
 Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;

```

QY 11 LPLLLLPAPPAKPPCHRCRGLVDKFNQGVDTAKKNGGNTAMEKTLSTKYES 70
DB 11 LLAALLLVLPVSEA-----NFLSKQASQVLVRRR--ANSLLLETKQGNLER 56
QY 71 SEIRLEILEIGCESSDPECNMLE--AOEHLLEAMWL-----QLKSEYP 113
DB 57 -----ECIEELCNKE--EAREVFENDPETDYFPRKYLIVCLRSFOTGLFTAROSTNAYP 108
QY 114 DLEFMCVTKLVKCSPGTYGPDCLACQGSQSPGSGNGHCS-GDGSROGDSRCRCHMGY 172
DB 109 DLRS--CVNAIPDQSP-----LPCNEDGYMSCKRG--KASTCTCKRGW 149
QY 173 QGPLC---TDCMDGYFSSLRNETHSICTACDESKTSGGLTNRDCGEGEYVWL--DEG 226
DB 150 QGEKCEFDINECKD-----PSNINGGCSQICDNTPG--SYHC-SCKNGFVLMISNKK 197
QY 227 ACVDVDECAEPPPCSAOFCNANGSYTCECDSSVCCTGEGNCKECLISGTA--RE 284
DB 198 DCKDVDECSLKPSIGTA--VCKNIPGDFEC-----ECPEGRYNLIK 237
QY 285 HQGCAADVDCSLAEKTCVKKNNENYNTPGSYVCV--PDGFETED-----ACVP 332
DB 238 SKSCEDIDCS--ENMCA---OLCVNYPGGTYCYCDGKKGKFLADQKSCYVSVCLP 290
  
```

RESULT 15
 US-08-628-747-3
 ; Sequence 3, Application US/08628747
 ; Patent No. 6169070
 ; GENERAL INFORMATION:

APPLICANT: Chen, Jian
APPLICANT: Godowski, Paul J.
APPLICANT: Hammonds, R. Glenn
APPLICANT: Mark, Melanie P.
APPLICANT: Mather, Jennie P.
APPLICANT: Li, Ronghao
TITLE OF INVENTION: RECEPTOR ACTIVATION BY GAS6
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: One DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: United States
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/628,747
FILING DATE: 17-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,253
FILING DATE: 10-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/438,861
FILING DATE: 10-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: P929P2PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-225-1994
TELEFAX: 650-952-9881
TELEX: 910-371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-628-747-3

Query Match 11.9%; Score 239; DB 4; Length 676;
Best Local Similarity 25.7%; Pred. No. 3,1e-11;
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 22;
QY 11 LPLLLLPPEAKKPTPCRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70
DB 11 LPLLLLPPEAKKPTPCRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70
QY 71 SEIRLEITLEGICSSDFECNOMLE-AOEHLLEAMWL-----QKSEYP 113
DB 57 -----ECIEELCKNE--EAREVFENDPETDYFKYKLVCLRSFOTGLFTARQSTNAYP 108
QY 114 DLFEWFCYKTLKVCSPGTYGPDCLACOGGSGRPCSGNGHCS-GDGSROGDSGCHMGY 172
DB 109 DLRS--CVNALIPDQSP-----LPCNEDGYMSCKDG--KASFTCTCKPKGW 149
QY 173 GGPLC-----TDCMDGYFSLRNETHSICTACDESKCTCSGLTNRDCCGEVGVWL--DEG 226
DB 150 GGEKEKPEFINECKD-----PSNNGGCSQICDNTPG--SYHC-SCKNGFVMSLKK 197
QY 227 ACVDVDECAEPPPCSAOFCKNANGSYTCCEBDCSSCVGCTGEGGNGCKECSGYA--RE 284
DB 198 DCKDVDECSLKRISICGTA-VCKNIPGDFEC-----ECPEGYRYNLK 237
QY 285 HQCADVDECSLAERTCVARKNENCYNTPGSYVCVC--PDGFEETED-----ACVP 332

DB 238 SKSCEDIDECs--ENMCA---QLCVNYPGTYCYCGKKGFLAQDQKSCSEVSVCLP 290
RESULT 16
US-08-402-253-3
Sequence 3, Application US/08402253
Patent No. 621142
GENERAL INFORMATION:
APPLICANT: Chen, Jian
APPLICANT: Hammonds, R. Glenn
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
TITLE OF INVENTION: RSE RECEPTOR ACTIVATION
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,253
FILING DATE: 10-MAR-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Wendy M. Lee
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 929
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-402-253-3

Query Match 11.9%; Score 239; DB 4; Length 676;
Best Local Similarity 25.7%; Pred. No. 3,1e-11;
Matches 92; Conservative 37; Mismatches 115; Indels 114; Gaps 25;
QY 11 LPLLLLPPEAKKPTPCRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70
DB 11 LPLLLLPPEAKKPTPCRCGLVDKFNQGNVDTAKKFGGNTAMEEKLTKSYES 70
QY 71 SEIRLEITLEGICSSDFECNOMLE-AOEHLLEAMWL-----QKSEYP 113
DB 57 -----ECIEELCKNE--EAREVFENDPETDYFKYKLVCLRSFOTGLFTARQSTNAYP 108
QY 114 DLFEWFCYKTLKVCSPGTYGPDCLACOGGSGRPCSGNGHCS-GDGSROGDSGCHMGY 172
DB 109 DLRS--CVNALIPDQSP-----LPCNEDGYMSCKDG--KASFTCTCKPKGW 149
QY 173 GGPLC-----TDCMDGYFSLRNETHSICTACDESKCTCSGLTNRDCCGEVGVWL--DEG 226
DB 150 GGEKEKPEFINECKD-----PSNNGGCSQICDNTPG--SYHC-SCKNGFVMSLKK 197
QY 227 ACVDVDECAEPPPCSAOFCKNANGSYTCCEBDCSSCVGCTGEGGNGCKECSGYA--RE 284
DB 198 DCKDVDECSLKRISICGTA-VCKNIPGDFEC-----ECPEGYRYNLK 237

QY 285 HGQCADVDECSLAEKTCVRKNENCYNPGSYVCY--PDGFEETED-----ACVP 332

Db 238 SKSCEDIDECS--ENMCA---QLCVNYPGGYTTCYCDGKKGFKLAQDQKSCSEVVSVCCLP 290

RESULT 17
US-08-443-866B-3

1 APPLICANT: Godowski, Paul J.
2
3 APPLICANT: Hammonds, R. Glen
4
5 APPLICANT: Mark, Melanie R.
6
7 TITLE OF INVENTION: RSE RECEPTOR ACTIVATION
8
9 NUMBER OF SEQUENCES: 22
10
11 CORRESPONDENCE ADDRESS:
12

ADDRESS: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA

1. COMPUTER READABLE FORM:
2. MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
3. COMPUTER: IBM PC compatible
4. OPERATING SYSTEM: PC-DOS/MS-DOS
5. SOFTWARE: WinPatIn (Genentech)
6. CURRENT APPLICATION DATA:
7. APPLICATION NUMBER: US/08/443,866B
8. FILING DATE: 31-May-1995

; PRIOR APPLICATION DATA: 08/402253
 ; APPLICATION NUMBER: 10-MAR-1995
 ; FILING DATE: 10-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:

NAME: Lee, Wendy M.
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: P0929D1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3

```

;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 676 amino acids
;
; TYPE: Amino Acid
;
; TOPOLOGY: Linear
;
US-08-443-866B-3

```

Query Match	11.98	Score 239	DB 4	Length 676
Best Local Similarity	25.78	Pred. No. 3.1e-11		
Matches 92	Conservative 37	Mismatches 115	Indels 114	Gaps 22

```

QY 11 LPLLELPPEAKRPTPCRGRGLVDKRNQGVADPAKCNFGGMATMEKTLKSYES 70
Dd 11 LILILILILILILILILILILILILILILILILILILILILILILILILILILIL 11
11 LLACLLLVLPVSEA-----NFLSKQAQSVLYRKRR--NANSLLEETQGNLIER 56
QY 71 SEIRLLELLEGLCESSDFECNQMLE--AQEHELEAWL-----OLKSEYP 113
Dd 57 -----EEIEELCKNE--EAREVEFENDEPETYFPKYLVLCLRSFOTGLFAARSTNAP 108
114 DLFEWFCYKTLKVCSSPGATYGPDLACGGSGSQRCSGNGHCS--DDSGRQDGGSCRCRMGY 172
QY 109 DLRS--CVNALPDDCSP-----LLECNEDGYMSCKDQ--KASFCTCKPKGW 148
Dd 173 QGPLIC-----TTCMGYFSLRNETHSTCTACDESKCTCSGLTNRDCEBCEYGAWL--DEG 228
150 QGCEKCEPDINCKRD-----PSNINGGCSQICDMPG--STHC--SCKKGFWALSNKK 197
QY 227 ACVDVDECAAEPPCCSAAQCFCKNANGSYTCECDSCSCVGTGEGSPGNCKEICTSYA--RE 284
11 LILILILILILILILILILILILILILILILILILILILILILILILILILILIL

```

[illegible]

RESULT 18
US-08-525-940-23
; Sequence 23, Application US/08525940
; Patent No. 5866351

```

1  APPLICANT:  Franzsoff, Alex
2
3  APPLICANT:  Miranda, Luis R.
4
5  APPLICANT:  Wolf, Joseph R.
6
7  TITLE OF INVENTION:  CD4 + T-LYMPHOCYTE PROTEASES AND GENES
8
9  TITLE OF INVENTION:  ENCODING SAID PROTEASES
10
11 NUMBER OF SEQUENCES:  25

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

1 COMPUTER READABLE FORM:
2
3 MEDIUM TYPE: Floppy disk
4
5 COMPUTER: IBM PC compatible
6
7 OPERATING SYSTEM: PC-DOS/MS-DOS
8
9 SOFTWARE: Patent In Release #1.0, Version #1.25
10
11 CURRENT APPLICATION DATA:
12
13 APPLICATION NUMBER: US/08/525,940
14
15 FILING DATE:
16
17 DISTRIBUTION: 51
18

```

? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 08/368,852
 ? FILING DATE: 01-JAN-1995
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 08/088,322
 ? FILING DATE: 07-JUL-1993
 ? ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 23:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-940-23

```

Query Match	11.8%	Score 236;	DB 2;	Length 799;
Best Local Similarity	25.8%	Pred. No. 6.3e-11;		
Matches 85; Conservative	40;	Mismatches 119;	Indels 86;	Gaps 17

```

QY 38 VDRKNQGVADPAKKNPFEGGNGATAMEEKLTSKYESEETILLETLEGLGCESSPFECNONHLEAO 97
Db 498 VERERYSHVEDPTDQETEDYAGP---CDPECESEV-----GCDGPGPDHNCDLH-- 544
QY 98 EEHLAAMWLQKSEYPDLEFMEFCYKTL-----KVC--GSP-----GTVGPDLAC 140
Db 545 -----YYKTKLKN-----TRIVYSCPPGHHADKAKRCKKACAPNCESSGSHGDQMSC 593
QY 141 QGG---SQRPSCSGNGHSCGDSGRG---DGSGR-----CHMGYGPPLCTCDMDGYFSGLRN 190
Db 594 KYGFLELNENFNSCYTHCP-DSTYDPTKKNLCRCSEMKCTCTEPHNHTECRDGL--SLQG 650
QY 191 ETHSI-----CTACDESCKTCGSLTNRDCGECEYGVNLDEGACVDVDECAAP 238

```

Db 651 SRCVSCEDGRYFNGOCOPCHRFCAATCAGAGADGCTNCTEGYFMEJDRCVQ----- 702
QY 239 PRCSAOFCNNA--NGSYTCECDSSCVGCTGEGPNCKECTSGYAREHQ-----CAD 290
Db 703 -SCSISTYFHDSSNGKSKCKDCISCLTCNGPGFKNCTSCPSGLDLDLGMCGAICKD 761
QY 291 VDECSLAEK---TCVRKNENCYNTPGSYVC 317
Db 762 ATEESMAEGGFCMLVKKNLQCRKVLQOLC 791

RESULT 19

US-08-976-838-23
; Sequence 23, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-976-838-23

Query Match 11.8%; Score 236; DB 2; Length 799;
Best Local Similarity 25.8%; Pred. No. 6.3e-11;
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKRNGQWVDTAKNFGGNTAMEEKLTSKESSEIRLETLBLEGLCSSDFECNMLEAQ 97
Db 498 VERERYSVEDPTDYGTEYDAGP---CDEPCESEV-----CCDGGPDPCHNDCLH-- 544
QY 98 EEHLAAMLOLQKSEYPLDFEWFVCVKT-----KVC---CSP-----GTGPPDCLAC 140
Db 545 -----YYKLKNN-----TRICVSSCPGPHYHADKRCRCACAPNCSCSFESHDDQKSC 593
QY 141 QGG---SQRPCSGNGHSGDSSROG--DGSCR-----CHMGYQGLPCTDCMDGYFSSLRN 190
Db 594 KYGYFLNETNISCYTHCP--DSYQDTKKNLCKRCSEKCTCTEERHNCTECRDGL--SLQG 650
QY 191 ETHSI-----CTACDESKCTCSGLTNDGCEGEVGVWLDGACVYDUECAAP 238
Db 651 SRCVSCEDGRYFNGOCOPCHRFCAATCAGAGADGCTNCTEGYFMEJDRCVQ----- 702

QY 239 PRCSAOFCNNA--NGSYTCECDSSCVGCTGEGPNCKECTSGYAREHQ-----CAD 290
Db 703 -SCSISTYFHDSSNGKSKCKDCISCLTCNGPGFKNCTSCPSGLDLDLGMCGAICKD 761
QY 291 VDECSLAEK---TCVRKNENCYNTPGSYVC 317
Db 762 ATEESMAEGGFCMLVKKNLQCRKVLQOLC 791

RESULT 20

US-08-525-940-21
; Sequence 21, Application US/08525940
; Patent No. 5866351
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; APPLICANT: Miranda, Luis R.
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
; TITLE OF INVENTION: ENCODING SAID PROTEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,940
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,852
; FILING DATE: 01-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,322
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-525-940-21

Query Match 11.8%; Score 236; DB 2; Length 881;
Best Local Similarity 25.8%; Pred. No. 7.1e-11;
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKRNGQWVDTAKNFGGNTAMEEKLTSKESSEIRLETLBLEGLCSSDFECNMLEAQ 97
Db 580 VERERYSVEDPTDYGTEYDAGP---CDEPCESEV-----CCDGGPDPCHNDCLH-- 626
QY 98 EEHLAAMLOLQKSEYPLDFEWFVCVKT-----KVC---CSP-----GTGPPDCLAC 140
Db 627 -----YYKLKNN-----TRICVSSCPGPHYHADKRCRCACAPNCSCSFESHDDQKSC 675
QY 141 QGG---SQRPCSGNGHSGDSSROG--DGSCR-----CHMGYQGLPCTDCMDGYFSSLRN 190
Db 676 KYGYFLNETNISCYTHCP--DSYQDTKKNLCKRCSEKCTCTEERHNCTECRDGL--SLQG 732

```

OY      191 EHNST-----CTACDESCKTCGTLNRRPGCECEVGMVLIDEGACVDDECAPR   238
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      733 SMCVSCECDGRFRFNQODPCPHRFATCATAGADGCINCTEGEYFMDGNCVO-----   784
                                     | : | : | : | : | : | : | : | : | : |
OY      239 PCCSAAQFCCKNA--NGSYTCEECDSVCYCCTGECPGNCKECISGYAREHQ-----CAD   290
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      785 -SCSITYFFDHSENGYSCKKRCIDSLCINCNGPFGFNCTSCPSSGYLLDGLMCOMGAIC   843
                                     | : | : | : | : | : | : | : | : | : |
OY      291 VDECSLAEK---TCVRKNENCYNPTGSGSYC   317
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      844 ATEESWABEGGCMLYKKNNLCORKYLOQC   873
                                     | : | : | : | : | : | : | : | : | : |

RESULT    21
US-08-976-838-21
; Sequence 21, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; OS-08-976-838-21

```

```

Query Match 236: DB 2: Length 881;
Best Local Similarity 25.8%; Pred. No. 7,le-11;
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17.

OY 38 VDFKQGNVDTAKKKNFGGNTAMEKTKSYSSLETLLEIGLCESSDFECNOMLEAO 97
Db 560 VERFRRYSREDEPTDYGTEGYAGP---CDPCESEV-----GCDGPGPDHCNDLH-- 626
OY 98 EHHLEAMWLQKSEKPEPDLFEKFCVTKL-----KVC--CSP-----GTGYPPCLAC 140
Db 637 -----YYKKLLNN-----TRICVSSCPGHYHADKKRCRCAKPNESCFCGSHGQCMSC 675
OY 141 OGG---SORPSCGNHGSGDGSROG--DSSCR-----CHMGOGPLCTOCMGYPSLPN 190
Db 676 KYGYTLNETNSCVYHCP--DSGYQDTKKGLCKKCSBNCCTCTCFHFNCTCRQGL--SLDG 732
OY 191 ETHSI-----CTACDESKTSGLTNRDGCCECVAGVLDEGACVDVDECAEP 238

```

Db 733 SRCVSCEDGXYFNGGQCCPCPHRCATCAGAGAGCINTCEGYFMEDGRGVQ-----794
 Oy 239 PGCAAOFCRKA--NGSYCECDSSCVCGGEGCPGKCEKCSIGYAREHQ-----CAD 260
 Db 785 -SCSISYFFHDSBNGKRSCKKCOISLTCNGPFFKNTCSGPSYLLDLDGMCQMGATCKD 84.3
 Oy 291 VDCESLAEK---TCVRKKNENYINPFSYVC 317
 Db 844 ATTESMAEGGFCMLYKKNKNIQCKRVLDQLC 873

```

RESULT 22
US-08-525-940-18
: Sequence 18, Application us/08525940
: Patent No. 5866351
:
: GENERAL INFORMATION:
: APPLICANT: Mirzusooff, Alex
: APPLICANT: Miranda, Luis R.
: APPLICANT: Wolf, Joseph R.
: TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
: TITLE OF INVENTION: ENCODING SAID PROTEASES
: NUMBER OF SEQUENCES: 25
:
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/525,940
: FILING DATE:
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/368,852
: FILING DATE: 01-JAN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/088,322
: FILING DATE: 07-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2848-11-C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 915 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-525-940-18

```

	Query Match	Similarity	11.8%	Score 236	DB 2%	Length 915
	Best Local	Similarity	25.8%	Pred. No. 7.4e-11		
	Matches	85	Conservative	40	Mismatches	119
					Indels	86
					Gaps	17
Qy	38	VDKFNQGVADPAKKFNGGNTAMEEKTLSKYESRILLETILEGLCSSDFECNOMLEAQ	97			
	11:1	11:1	11:1	11:1	11:1	
Db	614	VERRRYSNVEEDPTDDYGTEDYAGP----	CDPECSFY-----	GCDDGPPDHCDNCLH--		660
Qy	98	EEHLEAMWLKSEYEPDLFEWFCVKT-----	KVC--CSP-----	GRYGPDLCLAC	140	
	11:1	11:1	11:1	11:1	11:1	
Db	661	-----YYXKLKNN-----	TRICVSSCPGRHYHADKKRCRCAPNCSCFSGSHGQCMSC		709	
Qy	141	QGG---SQRPSCGNGHCSDGDSRGG--	DGSCR-----	CHMGYQGLCTCDMDGYFSSLRN	190	

Db 710 KYGFLENETSCVTHCP-DGSYODTKKLNCRKCSNCKTCTEFHNCYECRGL--SLQG 766
QY 191 ETHSI-----CTADESKTCSGLNRDCCGCEVGMVDEACVYVDECAEP 238
Db 767 SRCVSCEDGRYFNGQDQPCHRFCATCAGAGADGCINCITBGFMEDERCVY----- 818
QY 239 PCCSAOFCCKNA--NGSYTCEBDCSSVCGTGEFGNCKECSIGYAREHQ-----CAD 290
Db 819 -SCSISYTFDHSSENGYKSKCKDCISLTCNGPGRKNTSCPSGYLLDLGMCQGAICKD 877
QY 291 VDECSLAEK---TCVRKNENCYNTPGSYVC 317
Db 878 ATEESMAEGGFCMLVKKNLRCORVLOQLC 907

RESULT 23

US-08-976-838-18
Sequence 18, Application US/08976838
Patent No. 5981259
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-838-18

Query Match 11.8%; Score 236; DB 2; Length 915;
Best Local Similarity 25.8%; Pred. No. 7.4e-11;
Matches 85; Conservative 40; Mismatches 119; Indels 86; Gaps 17;

QY 38 VDKFNOGVADPAKKNFGGNTAMEEKTLSKYESIRLEILELEGCESSDSEFCNOMLEAQ 97
Db 614 VERRRYSVNEPTDDYGTEDYAGP-----CDPECSFV-----GDDGFGPDHCNCLAH-- 660
QY 98 EHLLEAMVLQKSEYPDLEFWEFCVKT-----KYC-CSP-----GTYPDCLAC 140
Db 661 -----YYKKLKN-----TRICVSSCPGHYHADKRCRCAPNCESCFGSHGQCMSC 709
QY 141 QCG---SQRPCSGNCHGSGDGRG--DGSCR-----CHMGVQGLCTDCMDGYRSSLRN 190
Db 710 KYGFLENETSCVTHCP-DGSYODTKKLNCRKCSNCKTCTEFHNCYECRGL--SLQG 766

QY 191 ETHSI-----CTADESKTCSGLNRDCCGCEVGMVDEACVYVDECAEP 238
Db 767 SRCVSCEDGRYFNGQDQPCHRFCATCAGAGADGCINCITBGFMEDERCVY----- 818
QY 239 PCCSAOFCCKNA--NGSYTCEBDCSSVCGTGEFGNCKECSIGYAREHQ-----CAD 290
Db 819 -SCSISYTFDHSSENGYKSKCKDCISLTCNGPGRKNTSCPSGYLLDLGMCQGAICKD 877
QY 291 VDECSLAEK---TCVRKNENCYNTPGSYVC 317
Db 878 ATEESMAEGGFCMLVKKNLRCORVLOQLC 907

RESULT 24

US-08-525-940-15
Sequence 15, Application US/08525940
Patent No. 586351
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
APPLICANT: Wolf, Joseph R.
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
TITLE OF INVENTION: ENCODING SAID PROTEASES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,940
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/368,852
FILING DATE: 01-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/088,322
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-940-15

Query Match 11.5%; Score 230.5; DB 2; Length 288;
Best Local Similarity 28.8%; Pred. No. 5.6e-11;
Matches 66; Conservative 31; Mismatches 81; Indels 51; Gaps 12;

QY 128 CSP-----GTYPDCLACG---SQRPCSGNCHGSGDGRG--DGSCR-CHMGY--- 172
Db 64 CAPNCESGFGSHGQCMSCRYGFLNETNSCVTHCP-DGSYODTKKLNCRKCSNCFRTC 122
QY 173 -QGPLCTDCMDGYFSSLRNETHSI-----CTADESKTCSGLNRDCCGCEV 219
Db 123 TEFHICTECRDGL--SLQSGRCSVSCEDGRYFIGDQCPCHRFATCAGAGADGCINCTE 180

QY 220 GWLIDGACVYDVECAEAEPFPCSAOFCCKNA--NGSYTCECDSCSVGCTGEGPGNCKEC 277
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 GYMEGRCVQI-----CSISYFHDHSSNGYKSKCKKDISCLTCNGPFGKNTSC 231
QY 278 ISGYAREHGO-----CADYDECSLAEK--TCVRKNENCTNPGSYVC 317
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 232 PSGYLLDLGCMGCAICKATEESMAEGFCMLVKKNNLQORRVLOOLC 280

RESULT 25

US-08-976-838-15
; Sequence 15, Application US/08976838
; Patent No. 5981259
; GENERAL INFORMATION:
; APPLICANT: Franzusoff, Alex
; TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASE NUCLEIC ACID
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,838
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2848-11-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-976-838-15

Query Match 11.5%; Score 230.5; DB 2; Length 288;
Best Local Similarity 28.8%; Pred. No. 5.6e-11;
Matches 66; Conservative 31; Mismatches 81; Indels 51; Gaps 12;

QY 128 CSP-----GYPPDLACOGG---SORPCSGNGHSGDSSROG--DGSRC-CHMGY-- 172
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 64 CAINCESCFESHGDCMCKCYGFFLNETNSCTVHCP--DSYODTKNNLCRKSENFKTC 122
QY 173 -QSPILCTDCMDGYFFSLRNTHSI-----CTADESCRTCSGLTNDGCEV 219
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 123 TERHICTECHDGL--SLQGSRCVSCEDGRYFTGQDCQPCRHFAFCAGAGADCLNCTE 180
QY 220 GWLIDGACVYDVECAEAEPFPCSAOFCCKNA--NGSYTCECDSCSVGCTGEGPGNCKEC 277
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 GYMEGRCVQI-----CSISYFHDHSSNGYKSKCKKDISCLTCNGPFGKNTSC 231
QY 278 ISGYAREHGO-----CADYDECSLAEK--TCVRKNENCTNPGSYVC 317
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 232 PSGYLLDLGCMGCAICKATEESMAEGFCMLVKKNNLQORRVLOOLC 280

RESULT 26

US-07-907-190-1
; Sequence 1, Application US/07907190
; Patent No. 5321123
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H.
; APPLICANT: Fernandez, Jose A.
; TITLE OF INVENTION: PROTEIN S POLYPEPTIDES AND ANTI-PEPTIDE
; TITLE OF INVENTION: ANTIBODIES THAT INHIBIT PROTEIN S BINDING TO CAB BINDING
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; STREET: 10666 No. 5321123th Torrey Pines Road, Mail Drop TPC8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/907,190
; FILING DATE: 19920701
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/724,746
; FILING DATE: 02-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Filling, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCRI119P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-07-907-190-1

Query Match 11.3%; Score 227.5; DB 1; Length 635;
Best Local Similarity 28.6%; Pred. No. 2.3e-10;
Matches 70; Conservative 26; Mismatches 74; Indels 75; Gaps 16;

QY 107 QLKSEYPDLEFEMPCVTKLKCSPGTYGPDCLACOGGSGRPSGNGHCS-GDSSROGDGS 165
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 QSTNATPDLRS--CVNAIPGCGSP-----LPCMEDGYMSCKDG--KASFT 101
QY 166 CRCHMGYGPPLC---TDCMDGYFFSLRNTHSICTACDESCRTCSGLTNRDCGCEVGM 221
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 102 CTCKPQMGCKECEFIDNECKD-----PSNINSGCSQICDNTPG--SYHC-SCKNFP 149
QY 222 VL--DEGACVYDVECAEAEPFPCSAOFCCKNANGSYTCECDSCSVGCTGEGPGNCKEC 279
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 150 VMLSNKKKDKDVDECSLKPSICGTA--VCKNIPEDFEC-----ECPF 189
QY 280 GYA--REHGQCADVDECSLAEKTCVRKNENCTNPGSYVC--PDGFETED----- 328
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 190 GYRYNLKSKSCBEDIDBS--ENNCA---QLCVAYPGGYTCYCGKKGFKLAQDQCKCEV 244
QY 329 -ACVP 332
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 245 SVCLP 249

RESULT 27

US-07-985-691-2

Sequence 2, Application US/07985691
Patent No. 5405946
GENERAL INFORMATION:
APPLICANT: Griffith, John H
APPLICANT: Bouma, Bonno N
APPLICANT: Bertina, Rogier
TITLE OF INVENTION: RECOMBINANT PROTEIN S VARIANTS DEFICIENT
IN CABP BINDING ACTIVITY, BUT HAVING APC COFACTOR ACTIVITY,
TITLE OF INVENTION: IN CABP BINDING ACTIVITY, COMPOSITIONS AND THERAPEUTIC
METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 10666 No. 5405946th Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/985,691
FILING DATE: 19921202
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSP0042P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-985-691-2

Query Match 11.3%; Score 227.5; DB 1; Length 635;
Best Local Similarity 28.6%; Pred. No. 2.3e-10;
Matches 70; Conservative 26; Mismatches 74; Indels 75; Gaps 16;
QY 107 QLKSEYPDLEFEMFCVKTLLKVCSPGTYGPPDCLACGGGSGRPGSGNGHCS-GDGSROGDGS 165
Db 61 QSTNAYPDLRS--CVNAIPDQCS-----LPCNEDGYMSCKDG--KASFT 101
QY 166 CRCHMGYOGPLC---TDCMGYFSSSLRNETHSICTACDESKTCSGLTNRDCGCEYGM 221
Db 102 CTCKRGMGGECEPDINCKD-----PSNINGGGSQICDNTPG--SYHC-SCKNGF 149
QY 222 VL--DEGACVDVDECAAEPPCSAAQFCFNANGSYTCECCSSCVGCTGEBGPNCKECIS 279
Db 150 VMLSNKKDKCKVDDECSLPSICGTA-YCKNIPGDEEC-----ECPE 189
QY 280 GYA--REHGOCADVDECSLAERTCVRNKNENYTPGYSVVC--PDGFEETD----- 328
Db 190 GYRNLSKSCSEDIDEC--ENMCA---QLCVNYPGTYCTCDGKKGKRLAODKSCREV 244
QY 329 -ACVP 332
Db 245 SVCLP 249

RESULT 28
US-08-436-804-2
Sequence 2, Application US/08436804
Patent No. 5656484
GENERAL INFORMATION:

APPLICANT: BOUMA, Bonno N.
APPLICANT: BERTINA, Rogier M.
TITLE OF INVENTION: PROTEIN S DELETION VARIANTS DEFICIENT IN
CABP BINDING ACTIVITY, BUT HAVING APC COFACTOR ACTIVITY,
TITLE OF INVENTION: CABP BINDING ACTIVITY, BUT HAVING APC COFACTOR ACTIVITY,
TITLE OF INVENTION: COMPOSITIONS AND THERAPEUTIC METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Young & Thompson
STREET: Second Floor, 745 South 23rd Street
CITY: Arlington
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,804
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/267,387
FILING DATE: 29-JUN-1994
APPLICATION NUMBER: EP 93 201 906.0
FILING DATE: 30-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: B038633
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-804-2

Query Match 11.3%; Score 227.5; DB 1; Length 635;
Best Local Similarity 28.6%; Pred. No. 2.3e-10;
Matches 70; Conservative 26; Mismatches 74; Indels 75; Gaps 16;
QY 107 QLKSEYPDLEFEMFCVKTLLKVCSPGTYGPPDCLACGGGSGRPGSGNGHCS-GDGSROGDGS 165
Db 61 QSTNAYPDLRS--CVNAIPDQCS-----LPCNEDGYMSCKDG--KASFT 101
QY 166 CRCHMGYOGPLC---TDCMGYFSSSLRNETHSICTACDESKTCSGLTNRDCGCEYGM 221
Db 102 CTCKRGMGGECEPDINCKD-----PSNINGGGSQICDNTPG--SYHC-SCKNGF 149
QY 222 VL--DEGACVDVDECAAEPPCSAAQFCFNANGSYTCECCSSCVGCTGEBGPNCKECIS 279
Db 150 VMLSNKKDKCKVDDECSLPSICGTA-YCKNIPGDEEC-----ECPE 189
QY 280 GYA--REHGOCADVDECSLAERTCVRNKNENYTPGYSVVC--PDGFEETD----- 328
Db 190 GYRNLSKSCSEDIDEC--ENMCA---QLCVNYPGTYCTCDGKKGKRLAODKSCREV 244
QY 329 -ACVP 332
Db 245 SVCLP 249

RESULT 29
US-08-267-387-2
Sequence 2, Application US/08267387

REFERENCE/DOCKET NUMBER: P-UW 2637
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1219 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-882-046-5

Query Match 11.2%; Score 224; DB 4; Length 1219;
Best Local Similarity 22.8%; Pred. No. 8.9e-10;
Matches 97; Conservative 42; Mismatches 101; Indels 186; Gaps 28;

QY 42 NOGAVD-----TAKNFGGNTAMEKTLKSYSSSEIRLEILEGLCESS--DFECNOM 93
DB 157 HSGMINSRQWOTLKQNTG-----IAHFE-YQIRV-----TCDDHYVGGCKNF 199
QY 94 LEAGEHLEAMWLOLKSYPDLF-EMPCVKTLKYCCSFTYGPDC--LACGGSSRPSG 150
DB 200 CRPRD-----DEFGHYACDQNGNKTCEGMWGPCCKAICRQG----CSP 240
QY 151 NGHSGSGSROGDSGRCHMGYQGPLCTDCM--DGYFSSLNNE-----191
DB 241 K-----HGSKLPDGCRCQYQWGLYCDKCIPIHGCYVGTGCTNERPQCLCTETNMGQLCDK 295
QY 192 -----THSIC-----TACDE-----SCK-TTC 206
DB 296 DLANYCTHQPCLNRTGTSNTGPKDYKQCSCEPGYSGPNCIEAHACLSDPCNHRGSKRTS 355
QY 207 SGLNRROGCEBENV-----LDE-----GAC-----V 229
DB 356 SGF---EC-ECSPGWTGPTCTNIDDCSPNNGSHGTCQDLVNGFKVCYCPQWTKTQL 411
QY 230 DVDECAAEPPPCSAOFCFNANGSYTC-----EECDSSCVGCTG--EGPGNCKE--- 276
DB 412 DANCEAK--PCVNARSKNLIASYCCLPGMGNQNDININDCGCCQNDASRDLYN 469
QY 277 -----CISGVAREHGQCADVECSLAETCYRKNECYNTPGSYVCVPGDF-----BET 326
DB 470 GYRCICPGYAGDHCE--RDIDEC--ASNPL-NGHCHCONEINRFQCLCTPFGSGLCQLD 525
QY 327 EDACVP 332
DB 526 IDYCEP 531

RESULT 32
US-08-368-852-15
Sequence 15, Application US/08368852
Patent No. 5691183
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
TITLE OF INVENTION: CD4+ T-LYMPHOCTE PROTEASES AND GENES
TITLE OF INVENTION: ENCODING SAID PROTEASES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/368, 852
FILING DATE: 05-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: 2848-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-368-852-15

Query Match 11.1%; Score 223.5; DB 1; Length 288;
Best Local Similarity 27.9%; Pred. No. 2e-10;
Matches 64; Conservative 28; Mismatches 86; Indels 51; Gaps 10;

QY 128 CSP-----GTYPPCLACOGG---SQRPCSGNGHSGDSROGDSGRCHMGYOG---- 174
DB 64 CAPNESCFSHGDCMCKGYFLNETNSGVTHCP-DGSYQDTKKMLCRKCSNXXTC 122
QY 175 ---PLCTCMGIFSSLANETHST-----CTACDESCCTGCGTLNRDCEBEV 219
DB 123 TEFHACTECRDGL--SLOGSRCSVCEDEGRYFGDCDCPCRXMTCAAGADGCINCTE 180
QY 220 GAVLDEGACVYVDECAAPPCSAOFCNA--NGSYCEBDCSSCVCTGEGPENCKE 277
DB 181 GYFMDGRCVOX-----CSISYIFDHSENGYKCKCKDLSCLTNGNPGFRNCTSC 231
QY 278 ISGVAREHQ-----CADVECSLAEK--TCVRKNECYNTPGSYVC 317
DB 232 PSGYLLDGMCGMGAICKDATEESWAEGBFCLVKKMLCQRKVLQQLC 280

RESULT 33
US-08-185-432-18
Sequence 18, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Busseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2523 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-185-432-18

Query Match 11.1%; Score 223; DB 1; Length 2523;
 Best local Similarity 23.5%; Pred. No. 2.4e-09;
 Matches 68; Conservative 20; Mismatches 101; Indels 100; Gaps 14;

QY 126 VC-CSPGYGPDCL---ACQGGSPRCSGNGHSGDGRGDSGRCHMGYOGPLCTD- 179
 DB 164 ICKCPGFGHATCKODINEC---SQNPKNKGQCINE---FGSYRCTQGRFGRNCDDE 217
 QY 180 -----CMDGYFSSLRNETHSICTA-----CDESKTC-----SGLTN 211
 DB 218 YPCNPSPCLNGGTCKQTDSTDTCTCLPEFSGQCEENIDCPNNCRNGTCVDGVT 277
 QY 212 RDCGEVGVNLDGACVYDDECAEPPPCSAQFCNANGSYTC-----EECDSS- 262
 DB 278 YNC-QCPPMWT-GQCTEDYDEQCLMPNACQNGCTCHNTYGGYNCVANGWGEDCSENI 335
 QY 263 --CVGCTGSEPGNCKE-----CISGYAREHGOC----- 288
 DB 336 DDCANACHSGATCHDRVASFYCECPHRTGLLCHLDNMCISNPKNCSNCDTNPVNGKA 395
 QY 289 -----ADVDECSLAETCYVRKNENCYNTPGSGYVCVCPDGF 323
 DB 396 ICTCPGTYGPACNNDYDECSLGANPC-ERHGCTNTLGSFGCNCFOGY 443

RESULT 34
 US-09-110-116-3
 ; Sequence 3, Application US/09110116
 ; Patent No. 6013479
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Hong
 ; APPLICANT: Coahan, Victoria L.
 ; APPLICANT: Stuart, Susan G.
 ; TITLE OF INVENTION: HUMAN EMRI-LIKE G PROTEIN COUPLED
 ; FILE REFERENCE: PF-0530 US
 ; CURRENT APPLICATION NUMBER: US/09/110,116
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 886
 ; TYPE: PRT
 ; ORGANISM: HOMO SAPIENS
 ; FEATURE:
 ; OTHER INFORMATION: 784994, GenBank
 US-09-110-116-3

Query Match 11.1%; Score 222.5; DB 3; Length 886;
 Best local Similarity 27.9%; Pred. No. 8.2e-10;
 Matches 79; Conservative 20; Mismatches 95; Indels 89; Gaps 17;
 QY 130 PGTYPDCLACQGGSPRCSGNGHSGDGRGDSGRCHMGYOGPLCTDGM--DGYES 186
 DB 74 PGRCKADIDEC--SQSPQPCGPNSSCK---NLSGRYKXCSLDGFSPTGNDWVPGKPGNFS 129
 QY 187 SLNETHSICTACDESKTCGSLTNRDCGE-----CEYGVNLDGACVVD 232
 DB 130 -----CTDINE---C-LTSRYCPRHSDCVNSMGSGSCQVGFISNSTCEDVN 174

QY 233 ECAEPPPCSAQFCNANGSYTC-----EECDSSC---V 264
 DB 175 EC-ADPRACEPHATCNNTGVNSCFNPGFESSGHLSCGGLKASCEDIDECTEMCPINS 233
 QY 265 GCTGSEPGN--CKECISGYAREHQ-----CADVDECSLAETCYVRKNENCYNTPG 313
 DB 234 TCTNNT-PGSYFC-TCHPGFAPSQGLNFTDQGYECDIDECRDQDSTC-CPNISCTNALG 290
 QY 314 SYVCVCPDGEETEDACVPAEATEGESPTQ---LPSREDL 353
 DB 291 SYSCGCTVGFH-----PNPESQKDNFSCORVLFKCEEDV 326

RESULT 35
 US-08-185-432-17
 ; Sequence 17, Application US/08185432
 ; Patent No. 5750652
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Bussau, Isabelle
 ; APPLICANT: Diederich, Robert J.
 ; APPLICANT: Xu, Tian
 ; APPLICANT: Matsuno, Kenji
 ; TITLE OF INVENTION: DELETED PROTEINS, NUCLEIC ACIDS, AND
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PENNIE & EDMONDS
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/185,432
 ; FILING DATE: 21-JAN-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-006
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2556 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 US-08-185-432-17

Query Match 11.1%; Score 222; DB 1; Length 2556;
 Best local Similarity 24.6%; Pred. No. 2.9e-09;
 Matches 75; Conservative 26; Mismatches 104; Indels 100; Gaps 15;
 QY 125 KYVCSPGTY-GPDCCLACQGGSPRCSGNGHSGDGRGDSGRCHMGYOGPLCTDMD 182
 DB 44 EACVCGGARVGPCCQDPNPLSTPCKNAGTCHVYDRGVADYACSCALDFSGPLCTLPD 103
 QY 183 GYESSLRNE-----THSICTACDESKTCGSLTNRDCGE-----CEYGVNLDGACV- 229
 DB 104 N--ACLTNCRNGCTODLLTFYKRCRCPGWSGKSCQADPCASIPCANNGGCLPFEAS 161
 QY 230 -----DVDECAEPPPCSAQFCNANGSYTC----- 256

```

Db 162 YICHCPSPFHGPTCQWQVNECGKPRLCRHGCTCHNEVGSYRCVCRATHHTGNCNCEMPYVP 221
QY 257 -----ECCDSSCV-GCTGEG-----PG-NCKE---CISGTARE 284
Db 222 CSPSPCONGTGCRPTGVTHEC--ACLPFTGQNCENIDDCPPNNCKNGACVGVNTY 279
QY 285 HGOC-----ADVDECSLAETCVRKNENCTNPGSYVCVCPDGF-----EETEDA 329
Db 280 NCPCPPEWTGQYCTEDYDECQLMFNAC-QNGCTCHNTHGGTNCVGVNWGEDCSENID 338
QY 330 CVPPA 334
Db 339 CASNA 343

```

```

RESULT 36
US-08-083-590A-20
; Sequence 20, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-083-590A-20

```

```

Query Match 11.1%; Score 222; DB 1; Length 2556;
Best Local Similarity 24.6%; Pred. No. 2.9e-09;
Matches 75; Conservative 26; Mismatches 104; Indels 100; Gaps 15;
QY 125 KVCSPPTV-GPDLACGGSQSPSCGNGHCS-GDGSNQGSGCRCHMGYOGPLCTDCMD 182
Db 44 EACVCGGAFVGPQCDNPCLSTPCKNAGTCHVYDRGVADYACSCALGFSGPLCLPLD 103
QY 183 GYFSSLRNE-----THSICTACDESCRKTCSGLTNRDGE---CEVGVWLDGACAV----- 229
Db 104 N--ACLTNPPCRNGSTCDLTLTTEYKCKRCPGWSKSCQADPCANPCANGGCLPPEAS 161
QY 230 -----DVDECAEPPEPCSAOFCCKNANGSYTC----- 256

```

```

Db 162 YICHCPSPFHGPTCQWQVNECGKPRLCRHGCTCHNEVGSYRCVCRATHHTGNCNCEMPYVP 221
QY 257 -----ECCDSSCV-GCTGEG-----PG-NCKE---CISGTARE 284
Db 222 CSPSPCONGTGCRPTGVTHEC--ACLPFTGQNCENIDDCPPNNCKNGACVGVNTY 279
QY 285 HGOC-----ADVDECSLAETCVRKNENCTNPGSYVCVCPDGF-----EETEDA 329
Db 280 NCPCPPEWTGQYCTEDYDECQLMFNAC-QNGCTCHNTHGGTNCVGVNWGEDCSENID 338
QY 330 CVPPA 334
Db 339 CASNA 343

```

```

RESULT 37
US-08-532-384-20
; Sequence 20, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-532-384-20

```

```

Query Match 11.1%; Score 222; DB 3; Length 2556;
Best Local Similarity 24.6%; Pred. No. 2.9e-09;
Matches 75; Conservative 26; Mismatches 104; Indels 100; Gaps 15;
QY 125 KVCSPPTV-GPDLACGGSQSPSCGNGHCS-GDGSNQGSGCRCHMGYOGPLCTDCMD 182
Db 44 EACVCGGAFVGPQCDNPCLSTPCKNAGTCHVYDRGVADYACSCALGFSGPLCLPLD 103
QY 183 GYFSSLRNE-----THSICTACDESCRKTCSGLTNRDGE---CEVGVWLDGACAV----- 229

```

Db 104 N-ACLTNRCNGSTCDLLTLEFYKCRCPBGWSGKSCQADPCASNPANGGCLPEFAS 161
 QY 230 -----DVDECALEPPPCSAAPFCCKNANGSYTC----- 256
 Db 162 YICHCPSPFHGPTCRDVADEGKRLCRHGGTCHNEVSGYRCVCAATHGNCERPYV 221
 QY 257 -----EECDSCV-GCTGEG-----PG-NCKE---CISGYARE 284
 Db 222 CSPSPCQNGGTCTPRTDVTYHEC--ACLPFTGQNCENENIDDCPGNNCKNGACVGVNTY 279
 QY 285 HGQC-----ADVDECSLAETKCVKRNENCMYTPGSYVCPCDGF-----EETEDA 329
 Db 280 NCPPEWTEGQCTEDVDYDEQLPMAC-QNGGTCHNTHGNGCVANGWTGEDCSENID 338
 QY 330 CVPPA 334
 Db 339 CASAA 343

RESULT 38

US-08-882-046-7
 ; Sequence 7, Application us/08882046
 ; Patent No. 6136952

; GENERAL INFORMATION:
 ; APPLICANT: LI, Linheng
 ; APPLICANT: Hood, Leroy
 ; APPLICANT: Krantz, Ian D.
 ; APPLICANT: Spinner, Nancy B.
 ; TITLE OF INVENTION: Human Jagged Polypeptide, Encoding
 ; TITLE OF INVENTION: Nucleic Acids and Methods of Use
 ; NUMBER OF SEQUENCES: 110
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/882,046
 ; FILING DATE: 25-JUN-1997
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; REFERENCE/DOCKET NUMBER: P-UW 2637
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1010 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-882-046-7

Query Match 11.0%; Score 220.5; DB 4; Length 1010;

Best Local Similarity 22.9%; Pred. No. 1.4e-09;

Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

QY 88 FEQNQMLEAOEHLLEAMWLQKSEYDLF-EMFCVTLKVCSPGTGGPDC--LAQGGG 144
 Db 194 FGCKKRCRPD-----DFGHHACDNGNKTCEGMWGPCNNAICRQG- 237
 QY 145 QRCPSGNGHCSGDSRGDSCHMGYQGPLCTDCM--DGYSSLRNE----- 191

Db 238 ---CSPK-----HGSKLPDRCQYGMGLYCDKCIPIHPCGVHIGINEPMOCLSETNKG 289
 QY 192 -----THSIC-----TACDE----- 201
 Db 290 GOLCDKDLNVCYGHQPCNLNGTCSNTGPKYQCSCEBGSFGPNCIEAHEACLSDPCHNNG 349
 QY 202 SKCTGSGLTNRDCEGEVGV-----LDE-----GAC----- 228
 Db 350 SKETS--LGFEFC-ECSPTGTPTCTSTNIDDCSPNNCSHGTCQDLVNGFKVCVCPQMTG 406
 QY 229 ----VDVDECALEPPPCSAAPFCCKNANGSYTC-----EECDSCVGTG--EGRGNC 274
 Db 407 KTCQDLANECBAK--PCVNAKSKCNLIASYICDCLPGMNGCNDINIINDCLQCCQNDASC 464
 QY 275 KE-----CISGYAREHGQCADVDECSLAETKCVKRNENCMYTPGSYVCPCDGF--- 323
 Db 465 RDLVNGYRCICPPGYAGDHC- RDIDEC--ASNPCL-NGHCHQNEINRRQCLPTGFSGN 520
 QY 324 --EETEDACVP 332
 Db 521 LCQDLIDYCEP 531

RESULT 39

US-09-068-740A-6
 ; Sequence 6, Application US/09068740A
 ; Patent No. 6337387

; GENERAL INFORMATION:
 ; APPLICANT: SAKANO, SEIJI
 ; APPLICANT: ITOH, AKIRA
 ; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
 ; FILE REFERENCE: KP-8447
 ; CURRENT APPLICATION NUMBER: US/09/068,740A
 ; CURRENT FILING DATE: 1998-06-18
 ; PRIOR APPLICATION NUMBER: JP 7-299611
 ; PRIOR FILING DATE: 1995-11-17
 ; PRIOR APPLICATION NUMBER: JP 7-311811
 ; PRIOR FILING DATE: 1995-11-30
 ; PRIOR APPLICATION NUMBER: PCT/JP96/03356
 ; PRIOR FILING DATE: 1996-11-15
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1036
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-068-740A-6

Query Match 11.0%; Score 220.5; DB 4; Length 1036;

Best Local Similarity 22.9%; Pred. No. 1.4e-09;

Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

QY 88 FEQNQMLEAOEHLLEAMWLQKSEYDLF-EMFCVTLKVCSPGTGGPDC--LAQGGG 144
 Db 163 FGCKKRCRPD-----DFGHHACDNGNKTCEGMWGPCNNAICRQG- 206
 QY 145 QRCPSGNGHCSGDSRGDSCHMGYQGPLCTDCM--DGYSSLRNE----- 191
 Db 207 ---CSPK-----HGSKLPDRCQYGMGLYCDKCIPIHPCGVHIGINEPMOCLSETNKG 258
 QY 192 -----THSIC-----TACDE----- 201
 Db 259 GOLCDKDLNVCYGHQPCNLNGTCSNTGPKYQCSCEBGSFGPNCIEAHEACLSDPCHNNG 318
 QY 202 SKCTGSGLTNRDCEGEVGV-----LDE-----GAC----- 228
 Db 319 SKETS--LGFEFC-ECSPTGTPTCTSTNIDDCSPNNCSHGTCQDLVNGFKVCVCPQMTG 375
 QY 229 ----VDVDECALEPPPCSAAPFCCKNANGSYTC-----EECDSCVGTG--EGRGNC 274
 Db 376 KTCQDLANECBAK--PCVNAKSKCNLIASYICDCLPGMNGCNDINIINDCLQCCQNDASC 433

```

Oy 275 KE-----CISYAEHGGCAVDDECSLAETKCYVRKKNENCYPTGSGYCVCPDGG--- 323
      ::      |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 434 RDVINGRYCICPPPGYAGDHCE-RDIDEC--ASNCL-NGGHCHQAEINRFQCLCPTEGSGN 489
      ::      |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Oy 324 --EETEDACVP 332
      ::      |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 490 LCOLDIDYCEP 500
      ::      |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

```

RESULT 40
US-09-068-740A-7
: Sequence 7, Application US/09068740A
: Patent No. 6337387
: GENERAL INFORMATION:
: APPLICANT: SAKANO, SEIJI
: APPLICANT: ITOH, AKIRA
: TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
: FILE REFERENCE: KP-8447
: CURRENT APPLICATION NUMBER: US/09/068,740A
: CURRENT FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: JP 7-259611
: PRIOR FILING DATE: 1995-11-17
: PRIOR APPLICATION NUMBER: JP 7-311811
: PRIOR FILING DATE: 1995-11-30
: PRIOR APPLICATION NUMBER: PCT/JP96/03356
: PRIOR FILING DATE: 1996-11-15
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 7
: LENGTH: 1187
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-068-740A-7

```

Query Match	11.0%	Score 220.5;	DB 4;	Length 1187;
Best Local Similarity	22.9%	Pred. No. 1.6e-09;		
Matches 85;	Conservative 33;	Mismatches 94;	Indels 159;	Gaps 22

Oy	88	FECONMTEAJOENLEAMWILQKSEYRDLF-EMPCVKTLKVOCSPTGYSPRC--LAOOGGS	144
Db	163	FGCNKFCPRD-----DFGHTACDDQNGKNTCEGMWPEPCNRAICRQG-	206
Oy	145	QRPSCGNHCSCGDSRGDSRCRMGYOGLTCDCM-DGYESSLNE-----	191
Db	207	---CSPK-----HSCSKLRGDCRCROYTQMGLYCDKCLPHRCVCHGICNEFMQCLCETNMG	258
Oy	192	-----THSIC-----TACDE-----	201
Db	259	GQLCDKDLNYCGTHQPCLNNGTCSNTPBPKYQSCSPREGSGSPNCIEAHACLSDPCHNRG	318
Oy	202	SCRKCSGLTINDCCECEYGVW-----LDE-----GAC-----	228
Db	319	SKETS--LGFEC-ECSPGWTGPRCSNIDDCSPNMSCSHGTCODLVNGFCYCPQMTG	379
Oy	229	---VDVECAEAPRPPCSAAQFCNANGSYTC-----EEOSSCVGCG--EGPGNC	274
Db	376	KTCQLDANECAK--PCYNAKSCKNLASYCDCLPGHMONGCNDININDLIGQCQNDASC	433
Oy	275	KE-----CISGYAREHGOACADVDECSLAETKCYVRKNENCYNTPGSYCVCPDGF--	323
Db	434	RDLVNGYRCICRPGYAGDHCE-RDIDEC--ASNPL-NGHCQACQEIINRFQCLCTPTGFSGN	489
Oy	324	---EETEDACVP	332
Db	490	LCQDLIDYCEP	500

```

RESULT 41
US-08-400-159-6
; Sequence 6, Application US/08400155
; Patent No. 5869282
; GENERAL INFORMATION:

```

1 APPLICANT: Ish-Horowitz, David
2 APPLICANT: Henriques, Domingos M. P.
3 APPLICANT: Lewis, Julian H.
4 APPLICANT: Myat, Anna M.
5 APPLICANT: Fleming, Robert J.
6 APPLICANT: Artavanis-Tsakonas, Spyridon
7 APPLICANT: Mann, Robert S.
8 APPLICANT: Gray, Grace E.
9 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
10 TITLE OF INVENTION: SERATE GENE AND METHODS BASED THEREON
11 NUMBER OF SEQUENCES: 20
12 CORRESPONDENCE ADDRESS:
13 ADDRESS: Pennie & Edmonds
14 STREET: 1155 Avenue of the Americas
15 CITY: New York
16 STATE: New York
17 COUNTRY: USA
18 ZIP: 10036-2711
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: floppy disk
21 COMPUTER: IBM PC compatible
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23 SOFTWARE: Patent Release #1.0, Version #1.30
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/400.159
26 FILING DATE: 07-MAR-1995
27 CLASSIFICATION: 435
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Mistrock, S. Leslie
30 REGISTRATION NUMBER: 18,872
31 REFERENCE/DOCKET NUMBER: 7326-029
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (212) 790-9090
34 TELEFAX: (212) 869-9741/8864
35 TELEX: 66141 PENNIE
36 INFORMATION FOR SEQ ID NO: 6:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 1218 amino acids
39 TYPE: amino acid
40 TOPOLOGY: linear
41 MOLECULE TYPE: protein
42 US-08-400-159-6

Query Match	11.0%;	Score 220.5;	DB 2;	Length 1218;
Best Local Similarity	22.9%;	Pred. No. 1.7e-09;		
Matches 85; Conservative	33;	Mismatches 94;	Indels 159;	Gaps 22;

OY	88	FECNOMLEAOEHNELAMWLDLXSEYPLDF-EMPEVKTLLKCCSPTGYGPRD--LAOCGGG	1444
Dd	194	FGCNFKFCRPR-----BFFGHAYACDDQGNKTCMBGMWGPRECNRAICROG-	237
OY	145	ORPCSGNHCSGSDGSROGDSCRCMHGYOGLCTIDCM--DGYSFSLRNE-----	191
Dd	238	---CSPK-----HSCRTLRGDCRQYRGQGLYCDKCTLRHECVAGTICNERPQCLETNNMG	289
OY	192	-----THSIC-----TACDE-----	201
Dd	290	GOLCDKDLNGCYTHQRPCLNGTGCSNTGPDKYQCSCPREYSRPNCEIABHACLSDPCNRG	349
OY	202	SKCTCGSLTLNRDCGECEYGVW-----LDE-----GAC-----	228
Dd	350	SCKETS--LGFEFC-BCSPGMTGPTCSTINDICSFNMSHGCTCODLVNGFYKCPCPOMTG	406
OY	229	--VDVECAAEPRPGSAAPFCNKANANSYTC-----BEDCSSCVGTG--EGRGNC	274
Dd	407	KTCODLANECEAK--PCVNNAASCKNNLASYYCDCLPGMGCNCDININDILGQOCNDASC	466
OY	275	KE-----CISGARERHGACAVUDECSLAEKTCVRKNENICYMTPGSYVCVPDGF---	323
Dd	465	RDLVNGRYHCICPPRYAGDHCE-RVIDEC--ASNPL--NGHCONEINRFQCLCFPTGFSGN	520
OY	324	--EETEDACVP	332

Db 521 LCQLDIDYCEP 531

RESULT 42

US-08-611-729A-6
; Sequence 6, Application US/08611729A
; Patent No. 6004924

GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David
APPLICANT: Henrique, Domingos M.P.

APPLICANT: Lewis, Julian H.

APPLICANT: Myat, Anna M.

APPLICANT: Fleming, Robert J.

APPLICANT: Artavanis-Tsakonas, Spyridon

APPLICANT: Mann, Robert S.

APPLICANT: Gray, Grace E.

TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE

TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESS: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/611,729A

FILING DATE: 06-MAR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-037

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 866-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1218 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-611-729A-6

Query Match 11.0%; Score 220.5; DB 3; Length 1218;
Best Local Similarity 22.9%; Pred. No. 1.7e-09;
Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

QY 88 FPCNOMLEAEOEHLBAMWLQKSEYDLE-EMFCVTKLVCCSPGTYPDC--LACQGS 144

DB 194 FPCNFCRRRD-----DFGHHACDQNGKTKMEGMMGECHRAICRG- 237

QY 145 QPCSGNGHCSGDGRGDSRCRHMGYOGPLCTDCM--DGYFSSLRNE----- 191

DB 238 ---CSPK---HGSKCLPGDCRCQYGMQGLYCDKCIHPGCVHIGICNBPWQCLCETNMG 289

QY 192 -----THSIC-----TACDE----- 201

DB 290 GOLCDKDLNYCGTHQPLCLNGTCSNTGPDKYQCSCEGYSGPNCEIAHACLSDPCHNMG 349

QY 202 SKTCSGLTNRDQGECEVGV-----LDE-----GAC----- 228

DB 350 SKETS--LGFEC--ECSPTGPTGTCSTINIDDCSPNNSHGHTCQDLYVNGFKVCYCPPTWGT 406

QY 229 ---VDVDECAEPPECASAOFCNNANGSYTC-----EECDSSVCGCTG--EGPNC 274

DB 407 KTCQDLANDCEAK--PCVNAKSKRLIASYYCDDLPGMMGQVCDINDINDCAGQCNDA 464

QY 275 KE-----CISGTAREHGQCADVDECSLAETKTVRKNNKYNTPGYSVYCPDGF--- 323

DB 465 RDLVNGYRCICPPGYAGDHCE-RDIDEC--ASNPL-NGGHQNEINRFQCLCPTGFSGN 520

QY 324 --EETEDACVP 332

DB 521 LCQLDIDYCEP 531

RESULT 43

US-08-882-046-2

; Sequence 2, Application US/08882046

; Patent No. 6136952

GENERAL INFORMATION:

APPLICANT: Li, Linheng

APPLICANT: Hood, Leroy

APPLICANT: Krantz, Ian D.

APPLICANT: Spinner, Nancy B.

TITLE OF INVENTION: Human Jagged polypeptide, Encoding

TITLE OF INVENTION: Nucleic Acids and Methods of Use

NUMBER OF SEQUENCES: 110

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/882,046

FILING DATE: 25-JUN-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-UW 2637

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1218 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-882-046-2

Query Match 11.0%; Score 220.5; DB 4; Length 1218;
Best Local Similarity 22.9%; Pred. No. 1.7e-09;
Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 22;

QY 88 FPCNOMLEAEOEHLBAMWLQKSEYDLE-EMFCVTKLVCCSPGTYPDC--LACQGS 144

DB 194 FPCNFCRRRD-----DFGHHACDQNGKTKMEGMMGECHRAICRG- 237

QY 145 QPCSGNGHCSGDGRGDSRCRHMGYOGPLCTDCM--DGYFSSLRNE----- 191

DB 238 ---CSPK---HGSKCLPGDCRCQYGMQGLYCDKCIHPGCVHIGICNBPWQCLCETNMG 289

QY 192 -----THSIC-----TACDE----- 201

DB 290 GOLCDKDLNYCGTHQPLCLNGTCSNTGPDKYQCSCEGYSGPNCEIAHACLSDPCHNMG 349

Page 23

```

: Patent No. 6337387
:
: GENERAL INFORMATION:
: APPLICANT: SAKANO, SEIJI
: APPLICANT: ITOH, AKIRA
: TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
: FILE REFERENCE: KP-8447
: CURRENT APPLICATION NUMBER: US/09/068,740A
: CURRENT FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: JP 7-299611
: PRIOR FILING DATE: 1995-11-17
: PRIOR APPLICATION NUMBER: JP 7-31811
: PRIOR FILING DATE: 1995-11-30
: PRIOR APPLICATION NUMBER: PCT/JP96/03356
: PRIOR FILING DATE: 1996-11-15
: NUMBER OF SEQ ID NOS: 48

```

```

; LENGTH: 1218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-068-740A-11

Query Match      11.0%; Score 220.5; DB 4; Length 1218;
Best Local Similarity 22.9%; Pred. No. 1.7e-09;
Matches 85; Conservative 33; Mismatches 94; Indels 159; Gaps 23;

OY 88 FECCNMLEAQEHLEAMWLLQLKSEYDLF-EWFFCYKTLKVCSPGTYPPDC--LACGGGS 144
    ||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 194 FGCNFFECRPD-----DFGHTACDONGNKTCTMBEWMPECCNRATCRGG- 237
    ||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

OY 145 QAPCSGNGCHSGDGSRGDGSRCRMGYGAPLCTDCM--DGYFSLRLNE----- 191
    || :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 238 ---CSRK-----HGSKRLPGDCRCQYGMQGLYCDKICIPHPGCVHIGICNEPMQCLCETNMG 289
    || :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

```

```

Db      230  GOLCDKDLNYCTGTHPCLNGTGTGCTGTCGDKYQCSCEGYSQPNCEIAEHACLSDPCHNRG  349
Oy      202  SCKTCSGLTNRDCCGEYGVW-----LDE-----GAC-----                228
           |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      350  SCKETS--LGEFC-BCSPGWTGPTCSNTIDDSPNNGSHGGTQDVLVNGFKVCPPQWTG  406
           |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Oy      229  ----VDVDECAAEPPPCSAOFCRKANASYC-----EECDSSCYAGCTG--EGPGNC  274
           :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      407  KTCQIDANECEAK--PCVNAANSCKKLLASYCDCLPGMGMGQNCIDINIDCLGQCONDNASC  464
           :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Oy      275  KE-----CISGAREHGOCADVDECSLAEKTCVRKNNENYMTPGSYVCVCPDGF---  323
           :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      465  RDLVNGRYACICPPRYAGHGCE-RDIDEC--ASNPL-NGGHQGNELNFPQCLCPFGFSGN  520
           :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Oy      324  --EETEDACVP  332
           :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      521  LCQDIDYCEP  531

RESULT  46
5258288-1
; Patent No. 5258288
; APPLICANT: Wydro Robert; Cohen Edward; Dackowski William
; Stefanlo John; Lundwall Ake; Dahlback Bjorn
; TITLE OF INVENTION: VECTOR CONTAINING DNA ENCODING MATURE
; HUMAN PROTEIN S
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/890.401
; FILING DATE: 25-JUL-1986
; SEQ ID NO.: 1:

```

Query Match 10.98; Score 218.5; DB 6; Length 675;

Best Local Similarity 28.8%; Pred. No. 1.3e-09;
Matches 69; Conservative 23; Mismatches 73; Indels 75; Gaps 16;

QY 112 YPLPFEMFCYKTLKAGCSPTGYPPDLACGGSGSOPCSGNGHCS-DDGSGDGSORCHM 170
DB 107 YPLPLRS-CVNAISDQCNP-----LPCNEDGFMTCKDG-QAFTTCLCKS 147
QY 171 GYOGPLC---TDCMDGYFSSLNETHSICTACDESKTCSGLJNRDCCGEVGMV--D 224
DB 148 GMGEGCESDINCKDPV-----NINGGCSQICENTPG--SYHC-SCKNGFVLSN 195
QY 225 EGACVYDECAAEPPPCSAQFCRNANGSTCECCSSCVGCTGEGPKNKECISGARE 284
DB 196 KPOCKVYDECVLRPSICGTA-VCKNIPGDEEC-----ECAEGKYN 235
QY 285 --HGACADVDECSLAETCYCRKNENCYNTPGSYVCV--PDGFEETED-----ACVP 332
DB 236 PVKSCDDVDEC--AENLCA---QLCVNTPGISTCTCDGKKFKLADQKSCAIVPCLP 290

RESULT 47

US-08-185-432-19
; Sequence 19, Application US/08185432
; Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Bussieu, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2703 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-19

Query Match 10.8%; Score 216.5; DB 1; Length 2703;
Best Local Similarity 23.3%; Pred. No. 8.3e-09;
Matches 97; Conservative 27; Mismatches 135; Indels 157; Gaps 20;

QY 41 FNGGAMVDTAKN-----FGGNTAWEEKTISKY-----ESSEIR 74
DB 130 FDSLECEIAVPNACDHVTCINGGTC--QLKTLBEYTCACANGYTGERCEFTKNICASSPCR 187

QY 75 LLEILEGLCESSDPECN-----QMLEAOEHLIEMWMLKSEYPLDFEMFCVK 122

DB 188 NGATCTALAGSSSFQSCPPGFTGDTCTSYDIEBCQSN-----PKYGGICVN 234

QY 123 T---LKVCCSPGYTPDC-----LACGGSORPCSGNGHCSGDSKRGDSCCHMYOCP 175

DB 235 THGSYOCMPPTGYTGDCCTKYPK--SPSPQONAGIGRSNGLSY---ECKCPKGFEEK 288

QY 176 LCTDMDGYFSSL-RNETHSICTACDESKTCSGLJNRDCCGEVGMVLDGACVYDVC 234

DB 289 NCEQNTDDCLGLHCQNGGTCIDGISDYTCRCPPNFTGRCOD-----DYDVC 335

QY 235 A-AEPPPCSAQFCRNANGSTCECCSSCV-GCTG---EBPGNKE-----CISG 280

DB 336 AQRDHPVCONGATCTNTHGSYC-----ICVNGMAGLDCSNNTDCKQAACFYATCIDG 390

QY 281 YAREHQCA-----VY 291

DB 391 VGSFYCQCTKTKTGLCHLDDACTSNPCHADACTSPINGSYACSCANGYGVDCSEDI 450

QY 292 DECSLAETCYCRKNENCYNTPGSYVCVCPDGFEETEDACVPAEATGES-PTQ 346

DB 451 DECD--QGSPECHNGICVNTPGSYRCNCSQGFPG-----PRCEININCESHPCQ 498

RESULT 48

US-08-185-432-16
; Sequence 16, Application US/08185432
; Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Bussieu, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-16

Query Match 10.7%; Score 215.5; DB 1; Length 2471;

MOLECULE TYPE: peptide

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-19

Query Match 10.7%; Score 215.5; DB 3; Length 2471;
Best Local Similarity 21.8%; Pred. No. 9e-09;
Matches 88; Conservative 43; Mismatches 110; Indels 163; Gaps 23;

QY 28 PIPCH--RCRGVDRKFNQGVNVDRAKNPFGGNTAMWEKILSKYES--SEIRLLEILEGI 82
DB 576 PDCPHHGQCQDDGIDSY-----TCICNPGYMAICSDQIDECTSSPCLNDGRCIDLVNG- 628
QY 83 CESSDEECNQMLEAQEHLFAMWLQKSEYPDLEFWFCVTKLVCCSPGTGYPDC-LACQ 141
DB 629 -----YQCN-----CQPGTSGVNCIEINFD 647
QY 142 GGSQRPCSGNGHCSGDSRGDSGRCHMGYOGPLC-----TDCMDGYFS 186
DB 648 DCASNPCI-HGICMDGINRY---SCVCSPGFTGRCNIDIDECAASNPCRKATCINGV-- 701
QY 187 SLRNETHSIC-----TACDESCKTC-----SGLTNRDCGECEYGVV-----LD 224
DB 702 ---NGFRICLPPEGPHHPSCISOVNECLSNPCINGNCTGLSGYKC-LCDAGWVGINCEVD 757
QY 225 EGAC-----VDVDECAAEPPPCSAAGFCRNANG 252
DB 758 KNECLSNPCQNGGTCNLVNGRYCTCKKGFKNQCVNIDECAASN--PCLNGSTCFDDIS 815
QY 253 SYTC-----ECCDSSCVGCTG---EGRGNKE-----CISGYAREHQCA-DVD 292
DB 816 GYTCHCVLPYTGKNCQTVLAPCSPNPECENAAVCKESPNEFSYSTCLCAPGWGGQRCITDID 875
QY 293 ECSIAERTCVRKNNQCYNFGSYVVCPPDGF-----EETEDACV 331
DB 876 EC--ISKPCM-NHGLCHNTGSGTYMCRCPPGFSGMDCEDIDIDL 916

Search completed: September 10, 2002, 11:08:46
Job time: 588 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 11:01:38 ; Search time 32.64 Seconds

(without alignments)
1201.258 Million cell updates/sec

Title: US-09-905-075-2
Perfect score: 2005
Sequence: 1 MRLPRRAAGLPLLLLP.....AEAEATGESPTQLPSREDL 353

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 65 summaries

1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2005	100.0	353	20 AAY08064
2	2005	100.0	353	20 AAY13344
3	2005	100.0	353	20 AAY05283
4	2005	100.0	353	21 AAB00169
5	2005	100.0	353	21 AAY83224
6	2005	100.0	353	22 AAY12319
7	2005	100.0	353	22 AAB68596
8	2005	100.0	353	22 AAB80212
9	2005	100.0	353	22 AAB61231
10	2005	100.0	353	22 AAB53075
11	1997	99.6	353	21 AAY88571

ALIGNMENTS

12	1887	94.1	329	22	AAB61233	Mature human TANGO
13	1716	85.6	318	21	AAB42711	Human OREF ORF2475
14	1647	82.1	353	21	AAY91870	Human apoptosis re
15	1573	78.5	386	22	AAU31870	Novel human secret
16	1481.5	73.9	348	22	AAB61240	Chinese hamster pr
17	972.5	48.5	420	22	AAB48106	Human TANGO 206 po
18	971.5	48.5	420	20	AAY13362	Amino acid sequenc
19	971.5	48.5	420	20	AAY05281	EGF-like homologue
20	971.5	48.5	420	21	AAB24396	Human PRO214 prote
21	971.5	48.5	420	21	AAB86569	Human PRO214 amino
22	971.5	48.5	420	22	AAH39899	Human polypeptide
23	971.5	48.5	420	22	AAU12316	Human PRO214 polyp
24	971.5	48.5	420	22	AAB68594	PRO214. Homo sapi
25	971.5	48.5	420	22	AAB80230	Human PRO214 prote
26	971.5	48.5	420	22	AAB48133	Human TANGO 206 va
27	971.5	48.5	420	22	AAB27228	Human EXMAD-6 SEQ
28	971.5	48.5	513	22	AAH41685	Human polypeptide
29	969.5	48.4	420	22	AAB48134	Human TANGO 206 va
30	969.5	48.4	420	22	AAB48135	Human TANGO 206 va
31	969	48.3	434	21	AAB38395	Human secreted pro
32	969	48.3	434	21	AAV76151	Human secreted pro
33	968.5	48.3	420	22	AAB48110	Mouse TANGO 206 po
34	967.5	48.3	420	22	AAB48137	Mouse TANGO 206 va
35	966.5	48.2	420	22	AAB48136	Mouse TANGO 206 va
36	965.5	48.2	420	22	AAB48138	Mouse TANGO 206 va
37	860.5	42.9	392	21	AAB38394	Human secreted pro
38	670	33.4	374	22	AAB56688	Drosophila melanog
39	463.5	23.1	242	21	AAV76114	Rat HT glycoprotein
40	463.5	23.1	242	21	AAB56053	Skin cell protein
41	406	20.2	140	22	AAB11766	Human apoptosis re
42	274.5	13.7	2189	11	AAH05222	Antigen GX5401FL e
43	266.5	13.3	637	21	AAV32346	Mouse cell surface
44	266.5	13.3	644	21	AAV79186	Haematopoietic ste
45	266.5	13.3	644	21	AAV79193	Haematopoietic ste
46	260	13.0	1121	22	AAV78887	Human fibrillin 3
47	256.5	12.8	1208	21	AAV70554	Human latent trans
48	256.5	12.8	1257	21	AAV70554	Human latent trans
49	256.5	12.8	1260	22	AAH1460	TANGO 275 mature p
50	256.5	12.8	1289	22	AAH61419	Human TANGO 275 p
51	255.5	12.7	1253	21	AAH12271	Mouse latent TGF-b
52	255.5	12.7	1253	21	AAH61420	Murine TANGO 275 p
53	254.5	12.7	1253	18	AAH27598	Human fibrillin type
54	254.5	12.7	566	12	AAH1148	Fibrulin A. Homo s
55	254.5	12.7	601	12	AAH1149	Fibrulin B. Homo s
56	254.5	12.7	601	18	AAH27599	Human fibrillin type
57	254.5	12.7	652	19	AAH49879	Amino acid sequenc
58	254.5	12.7	652	21	AAV32345	Human cell surface
59	254.5	12.7	683	18	AAH1150	Fibrulin C. Homo s
60	254.5	12.7	683	18	AAH27600	Human fibrillin type
61	254.5	12.7	686	22	ABG19385	Novel human diagn
62	254.5	12.7	703	18	AAH27601	Human fibrillin type
63	254.5	12.7	706	22	ABG19386	Novel human diagn
64	253	12.6	3680	22	ABH70878	Drosophila melanog
65	250	12.5	1833	16	AAH79478	Mouse LTRP-2. Mus

RESULT 1
ID AAY08064
AAV08064 standard; Protein: 353 AA.
11-SEP-2000 (first entry)
Human EGF-like homologue protein (PRO217) encoded by DNA32292 cDNA.
Inflammatory cell infiltration; immune response; T cell proliferation;
anti-inflammatory; anti-autoimmune; anti-diabetic; spondyloarthritis;
T cell-mediated disease; spondyloarthritis; sclerodermis; renal disease;
inflammatory myopathy; hemolytic anemia; thrombocytopenia; thyroiditis;
KW

KW diabetes mellitus; demyelinating polyneuropathy; Guillain-Barre syndrome;
 KW multiple sclerosis; polyneuropathy; hepatitis; cirrhosis; enteropathy;
 KW sclerosing cholangitis; inflammatory bowel disease; Whipple's disease;
 KW skin disease; dermatitis; psoriasis; asthma; allergic rhinitis; tumor;
 KW food hypersensitivity; urticaria; eosinophilic pneumonia; transplant;
 KW idiopathic pulmonary fibrosis; graft rejection; PRO245; human; PRO217;
 KW EGF-like.
 OS Homo sapiens.
 XX
 XX
 PN MO9914241-A2.
 PD 25-MAR-1999.
 XX
 XX
 PF 17-SEP-1998; 98WO-US19437.
 XX
 PR 17-SEP-1997; 97US-0059119.
 PR 18-SEP-1997; 97US-0059263.
 PR 28-OCT-1997; 97US-0063350.
 PR 12-NOV-1997; 97US-0065186.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066770.
 PR 04-JUN-1998; 98US-0088026.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Fong S, Goddard A, Gurney AL, Tumas D, Wood WI;
 XX
 DR WPI: 1999-229499/19.
 DR N-PSDB; AAX37671.
 XX
 PT Composition containing novel polypeptide PRO245, its agonist or
 PT antagonist -
 XX
 PS Example 1; Fig 6B; 177pp; English.
 XX
 CC This invention describes a novel composition containing (part from a
 CC carrier or excipient), a novel PRO245 polypeptide (I), its agonist or
 CC antagonist, or their fragments, for modulating: (i) infiltration of
 CC inflammatory cells into tissue; (ii) an immune response; or (iii) T cell
 CC proliferation. The composition increases or decreases any of the effects
 CC (i)-(iii). The products of the invention have anti-inflammatory,
 CC anti-autoimmune and anti-diabetic activity. (I), and its (ant)agonists
 CC and their fragments, are used to treat immune-related diseases,
 CC particularly T cell-mediated diseases. The diseases treated include
 CC systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic
 CC arthritis, spondyloarthropathies, systemic sclerosis (scleroderma),
 CC idiopathic inflammatory myopathies (dermatomyositis, polymyositis),
 CC Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune
 CC hemolytic anemia (immune pancytopenia, paroxysmal nocturnal
 CC hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic
 CC purpura immune-mediated thrombocytopenia), thyroiditis (Grave's disease,
 CC Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic
 CC thyroiditis), diabetes mellitus, immune-mediated renal disease
 CC (glomerulonephritis, tubulointerstitial nephritis), multiple sclerosis,
 CC idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome, chronic
 CC inflammatory demyelinating polyneuropathy, infectious hepatitis
 CC (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune
 CC chronic active hepatitis, primary biliary cirrhosis, granulomatous
 CC hepatitis, and sclerosing cholangitis, inflammatory bowel disease
 CC (ulcerative colitis; Crohn's disease), gluten-sensitive enteropathy, and
 CC Whipple's disease. Autoimmune or immune-mediated skin diseases including
 CC bullous skin diseases, erythema multiforme, contact dermatitis, psoriasis,
 CC asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity,
 CC urticaria, eosinophilic pneumonia, idiopathic pulmonary fibrosis,
 CC hypersensitivity pneumonitis, and transplantation associated diseases
 CC (graft rejection, and graft-versus-host-disease). (I), its (ant)agonists
 CC or fragment can also be used as an adjuvant in treatment of tumors.
 CC Antibodies against (I) can also be used for diagnosing such diseases.
 CC This sequence represents a human EGF-like homologue (PRO217) encoded by
 CC cDNA clone DNA32292 which is described in the invention.
 XX
 SQ Sequence 353 AA;

Query Match 100.0%; Score 2005; DB 20; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 353; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRLPRRALGLPLLLLPAPPAEAKKPPCHRCGLVDKFNQGVYDRAKKKNGGNTAM 60
 DB 1 mrlpraalglpllllpapaakpfcphrcrglvdkfngmvdtaakknfgyntaw 60
 QY 61 EEKTLISKYSSSEIRLEILEIGCESSDFECNQMLFAOEHELEAMWLOLKSEYPDLEFEMRC 120
 DB 61 eektliskyasseirllleileiglcessdfecnqmlaeqeehlawliqlkseypdllfwtc 120
 QY 121 VKTLKVCSPGTYGPDCLACQGGSPRCGNGCHGSDGSRQDGSRCRMHGQGPLCTDC 180
 DB 121 vktlkvcspgtygpdclacqggsqrpccngnchgsdgsrgdgsrctrmhgysgplctdc 180
 QY 181 MDGYFSSLLRNETHSICTACDESKTCSGLTNRDGCSEVGVWLDGACVYDVECAAEPP 240
 DB 181 mdgyfssllrnehstictacdesktcsgltnrdgcsevgwldgacvvdcaaepp 240
 QY 241 CSAAORCKRANGSYRCEECDSGVGCTGSGPGMKCKECISGYAREHGOCADVDECSIAEKT 300
 DB 241 csaaorckrangsyrcceedscvgctgsgpgmkckecisgyarehgcadvdscsiae 300
 QY 301 CVRKNNCYNTPGSYVCVCPDGFEEFEDACVPAEAATEGESPTQLPSREDL 353
 DB 301 cvrkncncyntpgsyvcvcpdgdfeetedacvpaeeategesptqlpsredl 353
 RESULT 2
 AAY13344
 ID AAY13344 standard; Protein: 353 AA.
 XX
 AC AAY13344:
 XX
 DT 25-JUN-1999 (first entry)
 XX
 DE Amino acid sequence of protein PRO211.
 XX
 KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophila areata;
 KW anti-thrombotic; wound healing; tissue repair.
 XX
 OS Homo sapiens.
 XX
 PN WO9914328-A2.
 XX
 PD 25-MAR-1999.
 XX
 PF 16-SEP-1998; 98WO-US19330.
 XX
 PR 25-NOV-1997; 97US-0066840.
 PR 17-SEP-1997; 97US-0059113.
 PR 17-SEP-1997; 97US-0059115.
 PR 17-SEP-1997; 97US-0059117.
 PR 17-SEP-1997; 97US-0059119.
 PR 17-SEP-1997; 97US-0059121.
 PR 17-SEP-1997; 97US-0059122.
 PR 17-SEP-1997; 97US-0059184.
 PR 18-SEP-1997; 97US-0059263.
 PR 18-SEP-1997; 97US-0059266.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 24-OCT-1997; 97US-0063486.
 PR 24-OCT-1997; 97US-0062814.
 PR 24-OCT-1997; 97US-0062816.

PR 24-OCT-1997; 97US-0063045.
 PR 24-OCT-1997; 97US-0063120.
 PR 24-OCT-1997; 97US-0063121.
 PR 24-OCT-1997; 97US-0063127.
 PR 24-OCT-1997; 97US-0063128.
 PR 27-OCT-1997; 97US-0063329.
 PR 27-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 28-OCT-1997; 97US-0063564.
 PR 29-OCT-1997; 97US-0063435.
 PR 29-OCT-1997; 97US-0063704.
 PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.

(GETH) GENENTECH INC.

PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;

DR WPI; 1999-229533/19.

DR N-PSDB; AAX52213.

PT New isolated human genes and polypeptides used in, e.g. treatment of
 XX gastrointestinal ulceration

PS Claim 12; Fig 2; 320pp: English.

XX AA13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to
 CC growth or survival of nerve cells including Parkinson's disease,
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophica areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.

XX Sequence 353 AA;

Query Match 100.0%; Score 2005; DB 20; Length 353;

Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRLPRRAALGLPULLLLPPAPEAAKPPCHRCGLVDKFNQGVDPATAKNGSGGTAW 60
 Db 1 mlrprraalglp11111ppapeaakppchrcrglvdkfngmvdltaknfggnltaw 60
 QY 61 EEKLSKYESESEIRLLEGLCESSDFECNOMLEAOEHLAWMLLKSFPPLFMPFC 120
 Db 61 eeklskysesseirlllelelelcesdfeecnmlaeehlaewmlkseypllfewfc 120
 QY 121 VKTLKVCSPGTYGPPDCLACGGGSRPCSGNGHSGSGSRGDSGCRHMGYQGPLCTDC 180
 Db 121 vktlkvcspgtygppdclacggsgprcsgngshsgsgsrqgdsgrchmgypplctdc 180
 QY 181 MDGYFSSLRNETHSICFACDESKCTGSLNRDCGECBVGWLDGACVDVDECAEAPP 240
 Db 181 mdgyfsslrnethsictacdesckltnrdcgeevgwldgacvddvdecaaeapp 240
 QY 241 CSAAOFCKNANGSYTCBECDSQVCGTGEFGNCKECISGARHGGCADVDECSLAEXT 300
 Db 241 csaaofcknangsytcsecdsqvcgctgefgnckecisgarhggcadvdecslext 300
 QY 301 CVRKNECYNTPGSGYVCPCDGFEEFEDACYPAAEATEGESPTQPSREDL 353
 Db 301 cvrkencyntpgsyvcpcdgfefeetadacypaaeatgesptqlpsredl 353
 RESULT 3
 AAY05283
 ID AAY05283 standard; Protein; 353 AA.
 AC AAY05283;
 XX 22-JUN-1999 (first entry)
 DE EGF-like homologue PRO211.
 KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246
 KW EBAF-2; Inhibitor; Tumour growth; cancer; EGF-like homologue;
 KW FGF-8 homologue.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO9914327-A2.
 PD 25-MAR-1999.
 PF 10-SEP-1998; 98WO-US18824.
 PR 25-NOV-1997; 97US-0066840.
 PR 17-SEP-1997; 97US-0059114.
 PR 17-SEP-1997; 97US-0059117.
 PR 18-SEP-1997; 97US-0059263.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 24-OCT-1997; 97US-0062816.
 PR 29-OCT-1997; 97US-0063704.
 PA (GETH) GENENTECH INC.
 PI Botstein D, Goddard A, Gurney A, Hillan K, Lawrence DA;
 PI Roy M, Wood WI;
 XX WPI; 1999-229533/19.
 DR N-PSDB; AAX28433.
 PT Antibodies against specific proteins overexpressed in tumours
 XX Example 1; Fig 17; 130pp: English.
 XX This sequence represents the EGF-like homologue PRO211.

PT Leukemia and for identifying compounds capable of inhibiting
 PT growth of neoplastic cells
 XX
 PS Claim 31: Fig 2, 133pp: English.
 XX
 CC Isolated PRO211, PRO228, PRO538, PRO172 or PRO182 polypeptides
 CC or their agonists (preferably anti-PRO agonist antibody or a small
 CC molecule mimicking the biological activity of PRO polypeptide) are
 CC useful in vitro or in vivo for inhibiting the growth of a tumor cell.
 CC Compositions comprising the PRO polypeptides are useful for
 CC inhibiting neoplastic cell growth and for treating cancer including
 CC breast, ovarian, renal, colorectal, uterine, prostate, lung, bladder,
 CC central nervous system cancer, melanoma and leukemia in a mammal.
 CC The PRO polypeptides are also useful for treating other disorders
 CC such as neuronal, glial, astrocytal, hypothalamic and other glandular,
 CC macrophagal, epithelial, stromal, blastocellic disorders and
 CC inflammatory, angiogenic and immunologic disorders as well as being
 CC useful for identifying agonists to PRO polypeptides by contacting the
 CC polypeptide with a candidate molecule and monitoring biological
 CC activity mediated by the polypeptide.
 CC
 XX
 SQ Sequence 353 AA:
 Query Match 100.0%; Score 2005; DB 21; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRLEPRALGLPLLLLPAPAEAKKPTPCRCGLVDKFNQGMVDTAKKNEGGNTAW 60
 DB 1 mrlpraalglpllllpapeaakpctchrcrglvdkfngmvdtkknfggnltaw 60
 QY 61 EEKLTSESESEHLELGLGESSDFECNOMLEAOEHLEAWWLKLSYPLTFMFC 120
 DB 61 eekltseeserilleglgessdfecnmleaeenleawwlqkseypltfmfc 120
 QY 121 VKTLKVCSPBTVGPDCIACOGSGSORPCSGNGHSCGDSRQDSCRCRHMVGQPLCTDC 180
 DB 121 vktlkvcspgtygpdclacogsgsrpcsgnghscgdsrgqdsgrcchmgypclctdc 180
 QY 121 VKTLKVCSPBTVGPDCIACOGSGSORPCSGNGHSCGDSRQDSCRCRHMVGQPLCTDC 180
 DB 121 vktlkvcspgtygpdclacogsgsrpcsgnghscgdsrgqdsgrcchmgypclctdc 180
 QY 181 MDGFFSLRNFTHSICRACDESKTCSGLTNRDCGECFVGWVLDGACVNDYDECAAEPP 240
 DB 181 mdgffslrnfthsicracescktcsgltnrdgcevgwvldgacvndydecaaepp 240
 QY 241 CSAAQFCNNNGSYTCBECSSCGCTGEGPNCKECISGARFREGCAVDDECSLAEXT 300
 DB 241 csaaqfcknangsytccecdscvycgcpqpnckecisgarfregcadvdecslaekt 300
 QY 301 CVRKENCYMTPGSYVCVCPDGFETEDACVPAEAETEGSEPTQLPSREDL 353
 DB 301 cvrkencyltpgsyvcvcpdgfeetedacvpaaeatetegseptqlpsredl 353

PT Modified-site 26..30
 FT /note- "cAMP and cGMP dependent protein kinase
 FT phosphorylation site"
 FT
 FT Modified-site 44..50
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 54..60
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 55..61
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 58..62
 FT /note- "Casein kinase II phosphorylation site"
 FT
 FT Modified-site 66..70
 FT /note- "Casein kinase phosphorylation site"
 FT
 FT Modified-site 81..87
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 86..90
 FT /note- "N-myristoylation site"
 FT
 FT Domain
 FT
 FT Modified-site 94..116
 FT /label- Leucine_zipper_pattern
 FT
 FT Modified-site 150..156
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 149..153
 FT /note- "Glycosaminoglycan attachment site"
 FT
 FT Modified-site 155..159
 FT /note- "Glycosaminoglycan attachment site"
 FT
 FT Modified-site 158..164
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 164..170
 FT /note- "N-myristoylation site"
 FT
 FT Domain
 FT
 FT Modified-site 166..178
 FT /label- EGF-like_domain
 FT
 FT Modified-site 190..194
 FT /note- "N-glycosylation site"
 FT
 FT Modified-site 197..201
 FT /note- "Casein kinase II phosphorylation site"
 FT
 FT Modified-site 210..214
 FT /note- "Casein kinase II phosphorylation site"
 FT
 FT Modified-site 251..255
 FT /note- "N-glycosylation site"
 FT
 FT Modified-site 252..258
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 255..259
 FT /note- "Casein kinase II phosphorylation site"
 FT
 FT Modified-site 295..299
 FT /label- Casein_kinase_II_phosphorylation_site
 FT
 FT Modified-site 308..320
 FT /note- "Asn and Asp hydroxylation site"
 FT
 FT Modified-site 313..319
 FT /note- "N-myristoylation site"
 FT
 FT Modified-site 339..343
 FT /note- "Casein kinase II phosphorylation site"
 FT
 FT Modified-site 349..353
 FT /note- "Casein kinase II phosphorylation site"
 FT
 XX
 XX WO200021996-A2.
 XX
 XX PD 20-Apr-2000.
 XX
 XX PE 05-Oct-1999; 99WO-US23089.
 XX
 XX PR 13-Oct-1998; 98US-0104080.
 XX
 XX PA (GENTECH) GENENTECH INC.
 XX
 XX PI Ashkenazi A, Goddard A, Gurney AL, Klein RD, Napier M, Wood WI,
 XX Yuan J;
 XX WPI: 2000-317943/27.
 XX DR N-PSDB: AA293700.
 XX
 PT Composition for inhibiting neoplastic cell growth and treating cancers
 PT of ovary, uterus, prostate, lung and bladder, comprises PRO211, PRO228,
 PT PRO538, PRO172 or PRO182 polypeptide or their agonist

XX Claim 5; Figure 2; 122pp; English.
 PS
 CC Compositions comprising a PRO211, PRO228, PRO538, PRO172 or PRO182
 CC polypeptide or their agonists, mixed with a carrier is useful for
 CC inhibiting neoplastic growth and treating tumors such as cancers of
 CC breast, ovary, renal, colorectal, uterus, prostate, lung, bladder,
 CC central nervous system, melanoma and leukaemia.
 XX
 SQ Sequence 353 AA:

Query Match 100.0%; Score 2005; DB 21; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRAAIGLLPLILLPAPAAKKPTPCHRCGLVDFKNGWMDTAKKNGGGNTAM 60
 DB 1 mrlpraaigllp11111pppeaakpchrclgvdfkngwmdtakkngggntaw 60
 QY 61 EKKTSKYESSSEIRLEILEGLCESSDFECNQMLEAQEHLAAMWLQKSEYDLEFEMFC 120
 DB 61 eektlskysesselrilleileglcessdfecngmlaeqehlawaqlkseydilefetc 120
 QY 121 VTKLVKCCSPGYTPDCLACGGGSGRPGCSGNHCSGDSGRQDGSRCRHMVGQPLCTDC 180
 DB 121 vtlkvccspgytpdclacggsgprpcsgnhsdgsrgdqgsrclmgyqgplctdc 180
 QY 181 MGVFSLNENHSTCTACDESKTCSGLTNRDCGCEVGMVLDGACVDVDECAEPP 240
 DB 181 mgyfslnelnhtctadescsktcsgltnrdcgcevgwvldgacvvdcaeaep 240
 QY 241 CSNAQPCNKANSGSYTCEDSCVCGTGEAPGNCKECISGYAREHQACADVDECSIAEKT 300
 DB 241 csaqpcnkansgsytcedscvctgeapgnckecisgyarehqacadvdecsiaekt 300
 QY 301 CVRKNECNTPTGSIYVCVCPDGEFETEDACVPAEAEATGESPTQLPSREDL 353
 DB 301 cvrkencntptgsiycvcpdgefetedacvpaeaeatgesptqlpsredl 353

RESULT 6
 AAU12319
 ID AAU12319 standard; Protein; 353 AA.
 AC AAU12319;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human PRO211 polypeptide sequence.
 XX
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200140466-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000WO-US32678.
 XX
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 09-DEC-1999; 99US-0170262.
 PR 16-DEC-1999; 99WO-US30905.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31243.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 XX
 PA (GEN) GENENTECH INC.
 PI Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 DR WPI: 2001-408281/43.
 DR N-PSDB; AAS21391.
 XX
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 12; Fig 296; 813pp; English.
 XX
 CC AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 353 AA:

Query Match 100.0%; Score 2005; DB 22; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRAAIGLLPLILLPAPAAKKPTPCHRCGLVDFKNGWMDTAKKNGGGNTAM 60
 DB 1 mrlpraaigllp11111pppeaakpchrclgvdfkngwmdtakkngggntaw 60
 QY 61 EKKTSKYESSSEIRLEILEGLCESSDFECNQMLEAQEHLAAMWLQKSEYDLEFEMFC 120
 DB 61 eektlskysesselrilleileglcessdfecngmlaeqehlawaqlkseydilefetc 120
 QY 121 VTKLVKCCSPGYTPDCLACGGGSGRPGCSGNHCSGDSGRQDGSRCRHMVGQPLCTDC 180
 DB 121 vtlkvccspgytpdclacggsgprpcsgnhsdgsrgdqgsrclmgyqgplctdc 180

Db 121 vktlkvcspgtygpdclacqgsqrsqngnhsqgsrgdgsccrmhgy99plctdc 180
 QY 181 MDGFSSLRNETHSICACDESCCTGSLTRDCECEVGVNLDEGACVVDDECAAEPP 240
 Db 181 mdgyfsslrnehtsictadescktsqgltrdcgecevgvnldegacvvddecaaepp 240
 QY 241 CSAQFCKNANGSYTCEDSSCGCTGEGPGNCKECISGYAREHGOCADVDECSLAET 300
 Db 241 csaaqfcknangsytccecdsscgctgepgnckecisgyarehgocadvdecslaekt 300
 QY 301 CVRKNENCYNTPGSYVCVCPDGFETEDACVPAPAEATEGESPTQLPSREDL 353
 Db 301 cvrkencyntpgsyvcvcpdgfeetedacvpaaategesptqlpsredl 353

RESULT 7
 AAB68596 standard; Protein: 353 AA.
 AC AAB68596:
 DT 27-APR-2001 (first entry)
 DE PRO211.
 KM Cytostatic; PRO protein; tumour; cancer.
 OS Homo sapiens.
 PN WO200105836-A1.
 PD 25-JAN-2001.
 XX 20-DEC-1999; 99WO-US30999.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28564.
 PA (GETH) GENENTECH INC.
 PI Botstein D, Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI;
 DR WPI: 2001-091968/10.
 DR N-PSDB: AAF60360.
 XX
 PT New antibody that binds to a PRO polypeptide, e.g. PRO187 and PRO533,
 PS useful for diagnosing and treating cancers -
 PS Claim 61: Fig 10; 196pp; English.
 CC The present invention relates to PRO proteins and coding sequences. The
 CC present sequence is one such PRO protein. It was found that the PRO genes
 CC are amplified in the genome of tumour cells. The gene amplification is
 CC expected to be associated with the overexpression of the gene product and
 CC contributes to tumorigenesis. Therefore, antagonists of PRO proteins are
 CC useful for the treatment of benign or malignant tumours, leukaemias,
 CC lymphoid malignancies and other disorders such as neuronal, glial,
 CC astrocytic, hypothalamic, glandular, epithelial, inflammatory and
 CC immunologic disorders.
 XX
 SQ Sequence 353 AA;

QY 1 MRLPRRALGLPLPLLLPPAPEAKKPTPCRHRCGLVDKPNOCADVDTAKKNGGNTAW 60
 Db 1 mrlprraaigllpllllppapeaakkptpcrhcrglvdkngmvdtaaknfggnbaw 60
 QY 61 EEKTLRSYSESEIRLELEBLCSSDFECNQMLEAEHLEAWMLDLKSEYPLDFEWC 120
 Db 61 eektlrsysselrlllellleglcsdfeecnqmlaeqehleawwldlksaypldfewfc 120
 QY 121 VKTLKVCSPGTYGPDCLACQGSQRPCSGNGHSGDGRDGSRCRHMCTGCPCLTDC 180
 Db 121 vktlkvcspgtygpdclacqgsqrsqngnhsqgsrgdgsccrmhgy99plctdc 180
 QY 181 MDGFSSLRNETHSICACDESCCTGSLTRDCECEVGVNLDEGACVVDDECAAEPP 240
 Db 181 mdgyfsslrnehtsictadescktsqgltrdcgecevgvnldegacvvddecaaepp 240
 QY 241 CSAQFCKNANGSYTCEDSSCGCTGEGPGNCKECISGYAREHGOCADVDECSLAET 300
 Db 241 csaaqfcknangsytccecdsscgctgepgnckecisgyarehgocadvdecslaekt 300
 QY 301 CVRKNENCYNTPGSYVCVCPDGFETEDACVPAPAEATEGESPTQLPSREDL 353
 Db 301 cvrkencyntpgsyvcvcpdgfeetedacvpaaategesptqlpsredl 353

RESULT 8
 AAB80212 standard; Protein: 353 AA.
 AC AAB80212:
 DT 24-APR-2001 (first entry)
 DE Human PRO211 protein.
 XX
 KW Human: PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
 KW antiparkinsonian nootropic; neuroprotective; vulnerrary; candid;
 KW antiangiogenic; vasotrophic; antiautomatic; antitumoural; cancer;
 KW antiarthritic; antidiabetic; antilivral; diabetes;
 KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
 KW ischaemia; inflammation.
 XX
 OS Homo sapiens.
 PN WO200104311-A1.
 PD 18-JAN-2001.
 XX
 PF 22-FEB-2000; 2000WO-US04414.
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 05-JAN-2000; 99WO-US00219.
 PA (GETH) GENENTECH INC.
 PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N,
 PI Filvaroff E, Fong S, Gao W, Garber H, Gerlitsen ME, Goddard A,
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavlin IJ,
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D,
 PI Williams PM, Wood WI;

XX WPI; 2001-081051/09.
 DR N-PSDB; AAF72371.
 XX
 PT Sixty one nucleic acids encoding PRO polypeptides which are useful in
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
 PT Alzheimer's disease) -
 XX
 PS Claim 1; Fig 2; 393pp; English.
 XX
 CC The present sequence is one of sixty one novel secreted and
 CC Transmembrane PRO polypeptides. The PRO polypeptides are
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
 CC endometrial bleeding, angiogenesis, ischaemia such as coronary
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
 CC diabetes and retinal disorders such as retinitis pigmentosa.
 CC The PRO nucleic acids have applications in molecular biology, including
 CC use as hybridization probes, and in chromosome and gene mapping.
 XX
 SQ Sequence 353 AA;

Query Match 100.0%; Score 2005; DB 22; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRRAALGLPLLLLPAPPAEAKKPPCHRCRGLVDKFNQGWVDTAKKNGGNTAM 60
 DB 1 mlprtraalglpllllpapeaakkpchrclrglvdkfngmwtdakknfggntaw 60
 QY 61 EKKTLKYSSEIRLLEIEGLCESSDFPCNOMLEAEOEHLAAMWLQKSEYPLDFEWC 120
 DB 61 eektlkysseirllleieglceessdfpcnmlleaeehlleawwlqlkseypldfewtc 120
 QY 121 VKTLKVCSPGTYGPDLCLACOGSGORPCSGNGHCSGDSROGSGRCHMGYGPCLCTDC 180
 DB 121 vktlkvcspgtygpdlclacogsgorpcsgngchscgdsrgsgscrchmgygpclctdc 180
 QY 181 MDGYFSLRNETHSICTACDESCKTCGSLTNRDGCCEVGNWLDGACVADVECAAEPP 240
 DB 181 mdgyfslrnethsictacdescctcsgltnrdgccevgwnwldgacvadvdecaaepp 240
 QY 241 CSAAOFCKNANGSYTCEEDSSCVGCTGEGPNCRCISGYAREHOCADVDECSLAEXT 300
 DB 241 csaafcknangsyteeedsscvgctgegpncrcisgyarehngcadvdecslaext 300
 QY 301 CVRKNNCYNTPGSGYVCVPDGFEEFEDACVPAEAATEGESPTQLPSREDL 353
 DB 301 cvrknnncyntpsgsyvcvpdgffeedacvpaeeategesptqlpsredl 353

RESULT 9
 AAB61231
 ID AAB61231 standard; Protein; 353 AA.
 XX
 AC AAB61231;
 XX
 DT 03-APR-2001 (first entry)
 XX
 DE Human TANGO 331 protein.
 XX
 KW Human; INTERCEPT 217; INTERCEPT 297; TANGO 276; TANGO 292; TANGO 325;
 KW TANGO 331; TANGO 332; cytosolic; antiinflammatory; antiarrhythmic;
 KW antiapoptotic; gene therapy; cancer; inflammatory disorder;
 KW cardiac disorder; arrhythmia; skin disorder; psoriasis.
 XX
 OS Homo sapiens.
 XX

PN 16-JUN-2000; 2000WO-US16658.
 XX
 PD 16-JUN-2001.
 XX
 PR 29-JUN-1999; 99US-0342364.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Kirst SJ, Holtzman DA, Fraser CC, Sharp JD, Barnes TS;
 XX
 DR WPI; 2001-061966/07.
 DR N-PSDB; AAF29457, AAF29458.
 XX
 PT Isolated human proteins are used for diagnosis, treatment and
 PT prevention of cancers, inflammatory disorders, cardiac disorders e.g.
 PT arrhythmia, and skin disorders e.g. psoriasis -
 XX
 PS Claim 8; Fig 6A-6C; 372pp; English.
 XX

The present sequence is given in a specification relating to isolated
 CC human proteins designated INTERCEPT 217, INTERCEPT 297, TANGO 276,
 CC TANGO 292, TANGO 325, TANGO 331 and TANGO 332. These proteins are useful
 CC as modulating agents or as targeting agents for developing agents to
 CC regulate cellular processes e.g. growth, proliferation, survival,
 CC differentiation and activity of human tissues. Diseases which can be
 CC diagnosed, prevented and treated by administration of these polypeptides;
 CC their nucleic acids and modulators include cancers, inflammatory
 CC disorders, cardiac disorders e.g. arrhythmia, and skin disorders e.g.
 CC psoriasis. Nucleic acids encoding the isolated proteins can be used to
 CC express the proteins in a host cell in gene therapy applications.
 CC Antisense molecules or ribozymes can be used to inhibit expression
 CC of the proteins in target cells. Fragments of the nucleic acid molecules
 CC encoding the isolated proteins are used as hybridisation probes and as
 CC polymerase chain reaction (PCR) primers.

Query Match 100.0%; Score 2005; DB 22; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.8e-133;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLPRRAALGLPLLLLPAPPAEAKKPPCHRCRGLVDKFNQGWVDTAKKNGGNTAM 60
 DB 1 mlprtraalglpllllpapeaakkpchrclrglvdkfngmwtdakknfggntaw 60
 QY 61 EKKTLKYSSEIRLLEIEGLCESSDFPCNOMLEAEOEHLAAMWLQKSEYPLDFEWC 120
 DB 61 eektlkysseirllleieglceessdfpcnmlleaeehlleawwlqlkseypldfewtc 120
 QY 121 VKTLKVCSPGTYGPDLCLACOGSGORPCSGNGHCSGDSROGSGRCHMGYGPCLCTDC 180
 DB 121 vktlkvcspgtygpdlclacogsgorpcsgngchscgdsrgsgscrchmgygpclctdc 180
 QY 181 MDGYFSLRNETHSICTACDESCKTCGSLTNRDGCCEVGNWLDGACVADVECAAEPP 240
 DB 181 mdgyfslrnethsictacdescctcsgltnrdgccevgwnwldgacvadvdecaaepp 240
 QY 241 CSAAOFCKNANGSYTCEEDSSCVGCTGEGPNCRCISGYAREHOCADVDECSLAEXT 300
 DB 241 csaafcknangsyteeedsscvgctgegpncrcisgyarehngcadvdecslaext 300
 QY 301 CVRKNNCYNTPGSGYVCVPDGFEEFEDACVPAEAATEGESPTQLPSREDL 353
 DB 301 cvrknnncyntpsgsyvcvpdgffeedacvpaeeategesptqlpsredl 353

RESULT 10
 AAB53075
 ID AAB53075 standard; Protein; 353 AA.
 XX

AC AAB53075;
 XX 28-FEB-2001 (first entry)
 XX
 DE Human angiogenesis-associated protein PRO211, SEQ ID NO:57.
 XX
 KW Human; angiogenesis-associated protein; PRO; endothelial cell growth;
 KW cardiac hypertrophy; cardiovascular disorder; endothelial disorder;
 KW angiogenic disorder; atherosclerosis; osteoporosis; hypertension;
 KW atherosclerotic infarction; diabetic retinopathy; rheumatoid arthritis;
 KW Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;
 KW Alzheimer's disease; Huntington's disease; stroke; drug screening;
 KW gene therapy; transgenic animal.
 XX
 OS Homo sapiens.
 XX
 PN NO200053753-A2.
 PD 14-SEP-2000.
 XX
 PF 05-JAN-2000; 2000MO-US00219.
 XX
 PR 08-MAR-1999; 99MO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 14-MAY-1999; 99US-0134287.
 PR 02-JUN-1999; 99MO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99MO-US20111.
 PR 08-SEP-1999; 99MO-US20594.
 PR 15-SEP-1999; 99MO-US21090.
 PR 15-SEP-1999; 99MO-US21547.
 PR 05-OCT-1999; 99MO-US23089.
 PR 30-NOV-1999; 99MO-US28313.
 PR 30-NOV-1999; 99MO-US28409.
 PR 02-DEC-1999; 99MO-US28564.
 PR 02-DEC-1999; 99MO-US28565.
 XX
 XX (GETH) GENENTECH INC.
 PA Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Goddard A;
 PI Godowski PJ, Gurney AL, Hillan KJ, Kuo SS, Mark MR, Marsters SA;
 PI Pisoni NF, Pittl RM, Watanabe CK, Williams PM, Wood WI;
 XX
 DR WPI: 2001-090793/10.
 DR N-PSDB: AAC97409.
 XX
 XX New isolated nucleic acid for producing a PRO polypeptide, analyzing
 PT genetic disorders and treating cardiovascular, endothelial or
 PT angiogenic disorders, such as atherosclerosis, wounds or cancer -
 XX
 XX Claim 69; Fig 24; 293pp; English.
 PS
 XX The invention relates to novel human angiogenesis-associated proteins
 CC designated PRO proteins (AAB53064-B53097), and to nucleic acids encoding
 CC PRO proteins. The invention also relates to vectors and host cells
 CC comprising a PRO nucleic acid, the recombinant production of a PRO
 CC protein, PRO antibodies specific for a PRO protein, fusion proteins
 CC comprising a PRO protein, agonists or antagonists of a PRO protein, and
 CC compounds which inhibit the expression of a PRO gene. The invention
 CC additionally encompasses methods of identifying modulators of PRO
 CC expression or activity; diagnosing a cardiovascular, endothelial or
 CC angiogenic disorder, or a susceptibility to such a disorder by detecting
 CC mutations in a PRO gene, or the expression level of a PRO gene within a
 CC particular tissue; treating a cardiovascular, endothelial or angiogenic
 CC disorder via the administration of a PRO protein, PRO nucleic acid, or
 CC PRO agonist or antagonist; a retroviral gene therapy vector comprising a
 CC PRO nucleic acid; and methods of inhibiting or stimulating endothelial
 CC cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the
 CC administration of a PRO protein, or an agonist or antagonist thereof.
 CC PRO nucleic acids, PRO proteins, antibodies against PRO proteins, PRO
 CC agonists and PRO antagonists may be used as therapeutic agents to treat

CC cardiovascular, endothelial or angiogenic disorders, such as
 CC atherosclerosis, osteoporosis, myocardial infarction, hypertension,
 CC diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis,
 CC endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's
 CC disease, or stroke. PRO nucleic acids are additionally useful in the
 CC recombinant production of PRO proteins, as hybridisation probes to
 CC screen libraries to isolate cDNAs with sequence identity to PRO proteins;
 CC to map genes encoding PRO proteins, to analyse genetic disorders, and in
 CC gene therapy. PRO nucleic acids can also be used to produce transgenic
 CC animals useful for the development and screening of potential
 CC therapeutic agents. The present sequence represents a PRO protein of the
 CC invention.
 CC
 XX
 XX Sequence 353 AA;
 XX
 SO
 Query Match 100.0%; Score 2005; DB 22; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1,8e-133;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRLPRRAALGILPLLLPPAPEAKKPTPOHRCGLVDKPNQGMVDTAKKNGCGNTAW 60
 Db 1 mrlprraalglpllllppapeaakptpchrrcrglvdklngmvtakkniggnltaw 60
 QY 61 EEKTLSEYSESEIRLEILEGLCESDFECNQMLEAEHELEAWMLQKSEYPLDFEWF 120
 Db 61 eektlseyseesirleileglcesdfecnmleaeheleawmlqksypldfewfc 120
 QY 121 VKTLKVCSPGTYGPDCLACGSGSRPCSGNGHSGDGSROGDGSCRCHMGYGPIC 180
 Db 121 vktlkvcspgtygpdclacgsgsrpcsgngshsgdgsrsgdgsrclmgygplctdc 180
 QY 181 MDGYESSLRNETHSICACDSCCKTCSGLTRDGCBEVGVILDEGACVYDDECAEAPP 240
 Db 181 mdgyesslrnethsicacdsccktcsgltnrdgceevgvildegacvddcaaeapp 240
 QY 241 CSAAPFCNMANGSYTCECDSYCVCTGEGPGNCKECISGARHGCACADVDECSLA 300
 Db 241 csaaqfcknangsytccecdssyvcvctggegpgnckecisgarhgcacadvdecslekt 300
 QY 301 CVRKNEHCYNTPGSYVCVCPDGFETEDACVPAEAETEGESPTOLPSREDL 353
 Db 301 cvrknehcyntpgsyvcvcpdgfetedacvpaeaetegesptqlpsredl 353
 RESULT 11
 AAY88571
 ID AAY88571 standard; Protein: 353 AA.
 AC AAY88571;
 XX
 XX 09-AUG-2000 (first entry)
 DE Human PRO211 amino acid sequence.
 XX
 KW Antibody: PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246
 KW PRO317; tumour growth inhibitor; cancer; diagnosis; treatment; human;
 KW cell growth; proliferation; epidermal growth factor; EGF; ADGFR;
 KW antibody dependent enzyme mediated prodrug therapy.
 OS Homo sapiens.
 XX
 PN WO200015666-A2.
 PD 23-MAR-2000.
 XX
 PF 08-SEP-1999; 99MO-US20594.
 PR 10-SEP-1998; 98US-0099803.
 PR 10-SEP-1998; 98MO-US18824.
 XX
 PA (GETH) GENENTECH INC.

PI Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI, Botstein D;
 XX WPI: 2000-271386/23.
 DR N-PSDB: AAA30040.
 XX
 PT New isolated antibodies which bind to specific polypeptides used for
 PT diagnosis and treatment of neoplastic cell growth and proliferation -
 XX
 PS Example 5; Fig 10; 200pp; English.
 CC This sequence represents a human PRO211 amino acid sequence. PRO211
 CC shares sequence homology with the epidermal growth factor protein
 CC sequence. The invention relates to isolated antibodies which bind to a
 CC polypeptide. The "PRO" polypeptides are encoded by genes which are over
 CC expressed in the genome of tumour cells. Vectors and host cells
 CC comprising the nucleic acid encoding the antibodies are used in the
 CC production of the antibodies. The antibodies and nucleic acids encoding
 CC them are used for diagnosing a tumour in a mammal. The antibodies are
 CC used for inhibiting the growth of tumour cells and identifying compounds
 CC that inhibit a biological or immunological activity of and/or expression
 CC of a PRO187, PRO533, PRO214, PRO240, PRO211, PRO261, PRO246 or
 CC PRO317 polypeptide. The antibody can be used in antibody dependent enzyme
 CC mediated prodrug therapy (ADEPT) by conjugating the antibody to a
 CC prodrug activating enzyme which converts a prodrug to an anti-cancer
 CC drug. The antibodies can be fluorescently labelled and monitored by light
 CC microscopy, flow cytometry or fluorimetry for diagnosis and prognosis of
 CC tumours.
 CC
 SQ Sequence 353 AA:

Query Match 99.6%; Score 1997; DB 21; Length 353;
 Best Local Similarity 99.7%; Pred. No. 6.5e-133;
 Matches 352; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRPRRAALGLPLLLLPAPRAAKRPCHRCGLVDFKNGAMDYAKKNGGGTAM 60
 DB 1 mrlpraaagllpllllpapaakrpchrcrglvdfkngmvdtaaknlggntaw 60
 QY 61 EEKTLKSYESSEIRLEIIEGLCESSDFECNQLLEAOEHLLEAMWLQKSEYDLEFEMFC 120
 DB 61 eektlkysesselrlelleiglcessdfecngmleaeheleawlqklsyepdlfemfc 120
 QY 121 VKTLKVCSSBGTYGPDCLACQGGSGRPGSGNGHSCGDSRQDSCGRCHMGYGGPLCTDC 180
 DB 121 vktlkvcssbgtygpdclacqgsggrpgsngshcsdgsrsgqscrcrlmgygplctdc 180
 QY 181 MDGFSLRNETHSITACDESCRKTCSGLTNRDGCCEVGMWVLDGACVVDVDECAEPP 240
 DB 181 mdgfslnrnhstictadescrktcsgltnrdgccevgmwvldgacvvdvdecaaepp 240
 QY 241 CSAAOECCKNANGSYTCEDDSQVCTGREGPNCCKECISGYAREHGOQADVDDECSLAEXT 300
 DB 241 csaaofcknangsyteeedssqvgctgregpncckecisgyarehgqcadvdecsiaekt 300
 QY 301 CVKKNENCNTPGSYVCVCPDGFETEDACVPPAEAEATGESPTOLPSREDL 353
 DB 301 cvrknenyntpysyvcvcpdgfeeteadacvpaaeaategesptqlpsredl 353

RESULT 12
 AAB61233
 ID AAB61233 standard; Protein; 329 AA.
 XX
 AC AAB61233;
 XX
 XX
 DT 03-APR-2001 (first entry)
 XX
 XX Mature human TANGO 331 protein.
 DE
 XX Human; INTERCEPT 217; INTERCEPT 297; TANGO 276; TANGO 292; TANGO 325;
 KW TANGO 331; TANGO 332; cytostatic; antiinflammatory; antiarrhythmic;
 KW antipsoriatic; gene therapy; cancer; inflammatory disorder;

KW cardiac disorder; arrhythmia; skin disorder; psoriasis.
 XX
 XX Homo sapiens.
 OS
 XX W020000638-027
 PN
 XX 1995JAN3001
 PD
 XX 16-JUN-2000; 2000WO-US16658.
 PF
 XX
 XX
 PR 29-JUN-1999; 99US-0342364.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Kirst SJ, Holtzman DA, Fraser CC, Sharp JD, Barnes TS;
 XX WPI: 2001-061966/07.
 DR
 XX Isolated human proteins are used for diagnosis, treatment and
 PT prevention of cancers, inflammatory disorders, cardiac disorders e.g.
 PT arrhythmia, and skin disorders e.g. psoriasis -
 PS Claim 8; Page 329-330; 372pp; English.
 CC
 CC The present sequence is given in a specification relating to isolated
 CC human proteins designated INTERCEPT 217, INTERCEPT 297, TANGO 276,
 CC TANGO 292, TANGO 325, TANGO 331 and TANGO 332. These proteins are useful
 CC as modulating agents or as targeting agents for developing agents to
 CC regulate cellular processes e.g. growth, proliferation, survival,
 CC differentiation and activity of human tissues. Diseases which can be
 CC diagnosed, prevented and treated by administration of these polypeptides
 CC their nucleic acids and modulators include cancers, inflammatory
 CC disorders, cardiac disorders e.g. arrhythmia, and skin disorders e.g.
 CC psoriasis. Nucleic acids encoding the isolated proteins can be used to
 CC express the proteins in a host cell in gene therapy applications.
 CC Antisense molecules or ribozymes can be used to inhibit expression
 CC of the proteins in target cells. Fragments of the nucleic acid molecules
 CC encoding the isolated proteins are used as hybridisation probes and as
 CC polymerase chain reaction (PCR) primers.
 CC
 SQ Sequence 329 AA:

Query Match 94.1%; Score 1887; DB 22; Length 329;
 Best Local Similarity 100.0%; Pred. No. 3.4e-125;
 Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AKRPCHRCRGVDFKNGAMDYAKKNGGGNTAMEEKLTSKYSSEIRLEIIEGLCE 84
 DB 1 akkrpchrcrglvdfkngmvdtaaknlggntaweektlkysesselrlelleiglc 60
 QY 85 SSDFECNQLLEAOEHLLEAMWLQKSEYDLEFEMFCVKTLYCCSPGYGPDCLACQGG 144
 DB 85 ssdfecngmleaeheleawlqklsyepdlfemfcvktlyccspgygpdclacqggs 120
 QY 145 QRPSCGNHSCGDSRQDSCGRCHMGYGGPLCTDCMDGFSSLRNETHSITACDESCR 204
 DB 145 qrpsgnshcsdgsrsgqscrcrlmgygplctdcmgdfsslnrnhstictadescr 180
 QY 205 TCSGLTNRDGCCEVGMWVLDGACVVDVDECAEPPCSAAOECCKNANGSYTCEDDSQV 264
 DB 205 tcsgltnrdgccevgmwvldgacvvdvdecaaeppcsaaofcknangsyteeedsscv 240
 QY 265 GCTGEGPNCCKECISGYAREHGOQADVDDECSLAEXTCVKKNENCNTPGSYVCVCPDGE 324
 DB 241 gctgегpncckecisgyarehgqcadvdecsiaektcvrknenyntpysyvcvcpdgfe 300
 QY 325 ETEDACVPPAEAEATGESPTOLPSREDL 353
 DB 301 eteadacvpaaeaategesptqlpsredl 329

RESULT 13

AAB42711
 ID AAB42711 standard; Protein; 318 AA.
 XX
 AC AAB42711;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human OREF ORF2475 polypeptide sequence SEQ ID NO:4950.
 XX
 KW Human: open reading frame: OREF; detection: cytosolic; hepatotropic;
 KW vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 KW
 OS Homo sapiens.
 XX
 PN MO200058473-A2;
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC76920.
 XX
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 4122-4123; 5507pp; English.
 XX
 CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human OREF open reading frames 1 to 3161. The OREF
 CC sequences have activities such as: cytosolic; hepatotropic; vulnerable;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antirheumatic; antihypertensive;
 CC antinaemic; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an OREF-associated disorder. The
 CC nucleic acids can be used to express OREF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
 CC coagulation, to inhibit thrombosis, and as a contraceptive.
 XX
 SQ Sequence 318 AA;

Query Match 85.6%; Score 1716; DB 21; Length 318;
 Best Local Similarity 97.4%; Pred. No. 3,5e-113;
 Matches 298; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 20 PAPAARAKPTPCRRCLGVYKFNQGNVDTAKKNGGNGNTAAEETKLSKESSEIRLLEIL 79
 11
 Db 2 paacpgyllapchrcrjlvdkfngmvdtkknfggntaweektlkysseirllleil 61
 QY 80 EGLCESSDFECNDMLEAQEEHLEAMWLQKSEYPDLEFMEFCVTKLKVCSPGTGPPCLA 139
 62 eglcessdfecnmleaeqeelhaewwllkseydplfewfvcktlkvcspgtyppcla 121
 Db
 QY 140 CQGSQRPCSGNGNCSGDSNRQGSRCRHMGVGRPLCTDCMCGYESSLKRREHSTTAC 199
 122 cqgsqrpcsgngncsgdsgsrqgdsrctmgygplctdcmgyfsslnethsttacc 181
 Db
 QY 200 DESCRTSGLTNRDCCGCEGVWVDEGACVDVDECAAPPCSAAPPCSKNANGSYTCEEC 259
 182 descrtsgltnrdccgeevgwldegacvddccaeappccsaagfcknangsytcceec 241
 Db
 QY 260 DSSCVGCTGEGPCNCKECISGYAREHGQCADVDECSLAERTCVKRNKNCYNTPGSYVCV 319
 242 dsscvgtggegpcnckecisgyarehgcadvedcslaektcvrknncyntpgsyvcv 301
 Db
 QY 320 PDGFEE 325
 302 pdgfee 307
 Db

RESULT 14
 ID AAY91870
 ID AAY91870 standard; Protein; 353 AA.
 XX
 AC AAY91870;
 XX
 DT 19-JUL-2000 (first entry)
 XX
 DE Human apoptosis related protein.
 XX
 KW Human apoptosis related protein; ARP; modulator; proliferative; cancer;
 KW programmed cell death; neurodegenerative; Alzheimer's Disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 140..330
 FT /label= EGF_EGF-1like_domain
 FT Domain 148..187
 FT /label= lamlnln_EGF-1like_domain
 FT Misc-difference 190
 FT /label= OTHER
 FT /note= "OTHER - not defined in the specification,
 encoded by GRA"

30-MAR-2000.
 24-SEP-1999; 99MO-US22270.
 24-SEP-1998; 98US-0160623.
 (MILL-) MILLENNIUM PHARM INC.
 Rhodadoust MM;
 WPI; 2000-283545/24.
 N-PSDB; AAA08503, AAA08504.
 New apoptosis related proteins and nucleic acid molecules used for
 regulating cellular processes e.g. programmed cell death -

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration,

Db 303 gpgnckqclsgyarengqcadvervpnxpektlceektktcyntpgssyvccvpdgireltr 304

CC disease, pulmonary heart disease, rheumatic fever, congenital heart
 CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
 CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
 CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune,
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents the human TANGO 206 polypeptide.

Sequence 420 AA:

Query Match 48.5%; Score 972.5; DB 22; Length 420;
 Best Local Similarity 48.2%; Pred. No. 1e-60;
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

QY 7 AALGLPLLLP-----PAEPA--KRPTRCRGLVDKFNQGNVDAKKKFGG 56
 DB 12 avlwglstflnlpplwlpqpsppspqphchtcrglyvdsfinkylertldnfgg 71
 QY 57 NTAMEKRTSKYESRIRLEILEGICSSDFECNOMLEAOEPHLEAWMLQKSEYDLE 116
 DB 72 ntaweenlskylkdsctrlvewlegvcsksdfecmlllseelweswfnhxygeapdlf 131
 QY 117 EMFCVKTLKVCSPGTTPGPDCLACGGSGRPGSGNGHSCGDSROGDSRCRMGYQPL 176
 DB 132 gwlcsdsklkcpgagtfpgscplpcpgyterpcgyqcegeglrgsgnhdccqaygga 191
 QY 177 CNDMDGYSSIRNETHSICACDSCSKTCGSLTNRDCECEVGNWLDGACVNDVECA 236
 DB 192 cggcglgyfleaenashlvcsacifgpcarsgpeesnciqckkqvalnhlkcvdldcgt 251
 QY 237 EPPPCSAQFCNANGSYTCECDSCVGTGEGPCNCKECSIGYAREHGOCADVDECSL 296
 DB 252 egangcagdfcvtntegsyecicakacjcmgagpgrckcspgyqyskcldvdece- 310
 QY 297 AKTCYRKKNENYNTPGSYVCVPGDFETEDACVP---PAEA----EATGE 342
 DB 311 -levcpgeknkqenteggyrciaegykmeglcvkqeglpesagffsentede 362

RESULT 18

AA13362 ID AAY13362 standard; Protein; 420 AA.

XX AAY13362;

XX 25-JUN-1999 (first entry)

XX Amino acid sequence of protein PRO214.

XX Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; AIDS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.

OS Homo sapiens.

XX MO9914328-A2.

XX 25-MAR-1999.

XX 16-SEP-1998; 98MO-US13330.

XX 25-NOV-1997; 97US-0066840.

XX 17-SEP-1997; 97US-0059113.

XX 17-SEP-1997; 97US-0059115.

XX 17-SEP-1997; 97US-0059117.

XX 17-SEP-1997; 97US-0059119.

XX 17-SEP-1997; 97US-0059121.

PR 17-SEP-1997; 97US-0059122.
 PR 17-SEP-1997; 97US-0059184.
 PR 18-SEP-1997; 97US-0059263.
 PR 18-SEP-1997; 97US-0059266.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 21-OCT-1997; 97US-0063486.
 PR 24-OCT-1997; 97US-0062814.
 PR 24-OCT-1997; 97US-0063045.
 PR 24-OCT-1997; 97US-0063120.
 PR 24-OCT-1997; 97US-0063121.
 PR 24-OCT-1997; 97US-0063127.
 PR 24-OCT-1997; 97US-0063128.
 PR 27-OCT-1997; 97US-0063329.
 PR 27-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 28-OCT-1997; 97US-0063564.
 PR 28-OCT-1997; 97US-0063565.
 PR 29-OCT-1997; 97US-0063704.
 PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065893.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.

(GETH) GENENTECH INC.

XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;

XX WPI; 1999-229533/19.

XX N-PSDB; AAX52233.

XX New isolated human genes and polypeptides used in, e.g. treatment of

XX gastrointestinal ulceration

XX Claim 12; Fig 40; 320pp; English.

XX AAY13344-403 represent secreted and transmembrane human proteins.

XX The cDNA sequences are obtained from cDNA libraries, prepared from

XX fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.

XX The encoded polypeptides have specific uses based on their homology to

XX known polypeptides, e.g. PRO211 and PRO217 can be used for disorders

XX associated with the preservation and maintenance of gastrointestinal

XX mucosa and the repair of acute and chronic mucosal lesions

XX (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal

XX ulceration and congenital microvillus atrophy), skin diseases associated

XX with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial

XX

100

PD 08-JUN-2000.

PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 09-DEC-1999; 99US-0170262.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 30-DEC-1999; 99WO-US30999.
 PR 06-JAN-2000; 2000MO-US00377.
 PR 06-JAN-2000; 2000MO-US00376.
 PR 11-FEB-2000; 2000MO-US03565.
 PR 18-FEB-2000; 2000MO-US04341.
 PR 18-FEB-2000; 2000MO-US04342.
 PR 22-FEB-2000; 2000MO-US04414.
 PR 24-FEB-2000; 2000MO-US04914.
 PR 24-FEB-2000; 2000MO-US05004.
 PR 01-MAR-2000; 2000MO-US05601.
 PR 20-MAR-2000; 2000MO-US07377.
 PR 21-MAR-2000; 2000MO-US07532.
 PR 30-MAR-2000; 2000MO-US08439.
 PR 17-MAY-2000; 2000MO-US13705.
 PR 22-MAY-2000; 2000MO-US14042.
 PR 30-MAY-2000; 2000MO-US14941.
 PR 02-JUN-2000; 2000MO-US15264.
 PR 10-NOV-2000; 2000MO-US30873.

(GETH) GENENTECH INC.

XX Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerltsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX

DR WPI: 2001-408281/43.
 N-PSDB: AAS21388.

XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical

PS Claim 12; Fig 290; 813pp; English.

XX AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.

XX Sequence 420 AA:

Query Match 48.5%; Score 971.5; DB 22; Length 420;
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

QY 7 AALGLPLLLLP-----PAPEAA--KKPLPCHRCGLVDKFNQGNVDTAKKNFGGG 56

Db 12 avlwgslflnlpplwlpqppqssppqpchpctcrglydstfngkterclrdnfgg 71
 QY 57 NNAMEEKLSTKYESSIRLEILEEGCESDDEFCNONMLEAOEHLEAMWLQKSEYPDF 116
 Db 72 ntaweeenlskydsctrlvevlyevcskdsftechlllseeivfswfthqgeapdlf 131
 QY 117 EMECVKTLKVCSPGTYPGDCIACOGSGRSCGNGHCSGDSROGDSRCRHMVGPL 176
 Db 132 gylcsdsllkccpagtfgpsclpqpgrterpcgygqcegeglrgsgnhdccagyggea 191
 QY 177 CTDCMDGFRSSLRNETHSICTACDESKTCSCGTRNDGCECEVGNWLDGACVYDDECA 236
 Db 192 cgcqgilyfteaerfnashlyvsacifgpacrcspseencilqckkyaalhnlkcvdidecgt 251
 QY 237 EPPPCSAOFCNNANSGYCECDSSVCCTGEGPCNCKECSGYAREHGOCADYDECSL 296
 Db 252 egancgaadqfvcnltsgysrcdcaakacigcmagaprrckkcsppyyqgyskcldvdece- 310
 QY 297 AKETCVKRNKNCYNTPGSYVCYCPDGFETEDACVP---PABA---EATEGE 342
 Db 311 -tevcpngkngenteggyrcicaegykqmeglcvkeqjpesagfsemtede 362

RESULT 24
 AAB68594
 ID AAB68594 standard; Protein; 420 AA.

XX AAB68594;

XX 27-APR-2001 (first entry)

XX PRO214.

XX Cytostatic; PRO protein; tumour; cancer.

XX Homo sapiens.

XX WO200105836-A1.

XX 25-JAN-2001.

XX 20-DEC-1999; 99WO-US30999.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 08-SEP-1999; 99WO-US20594.

XX 13-SEP-1999; 99WO-US20944.

XX 15-SEP-1999; 99WO-US21090.

XX 05-OCT-1999; 99WO-US23089.

XX 29-NOV-1999; 99WO-US28214.

XX 30-NOV-1999; 99WO-US28313.

XX 02-DEC-1999; 99WO-US28564.

XX (GETH) GENENTECH INC.

XX Botstein D, Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI;

XX WPI: 2001-091968/10.

XX N-PSDB: AAF60352.

XX New antibody that binds to a PRO polypeptide, e.g. PRO187 and PRO533,
 PT useful for diagnosing and treating cancers -

XX Claim 61; Fig 6; 196pp; English.
 CC The present invention relates to PRO proteins and coding sequences. The
 CC present sequence is one such PRO protein. It was found that the PRO genes
 CC are amplified in the genome of tumour cells. The gene amplification is
 CC expected to be associated with the overexpression of the gene product and
 CC contributes to tumourigenesis. Therefore, antagonists of PRO proteins are
 CC useful for the treatment of benign or malignant tumours, leukemias,
 CC lymphoid malignancies and other disorders such as neuronal, glial,

CC astrocytal, hypothalamic, glandular, epithelial, inflammatory and
 CC immunologic disorders.
 XX
 XX
 SQ Sequence 420 AA:

Query Match 48.5%; Score 971.5; DB 22; Length 420;
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

OY 7 AALGLPLLLLP-----PAPEAA--KKPPCHRCGLVDKFNQGVDTAKKNFGG 56
 DB 12 avlwgslflnlpplwlpqpsppppqphchrcglvdsfngkjlertlrdfnfgg 71
 OY 57 NTAMEKTLKYSSEIRLELEIGLCESSDFECNQMLEAQEHLAMWLQKSEYDLE 116
 DB 72 ntaweenlskydseclrvlelvgyvcsksdfechrlllelseelvswwlhkqgeapdlf 131
 OY 117 EMECVTLKVCSPGTYGPDCLACOGSGORPCSGNGHSGDGSROGSGSCRMHGYOGL 176
 DB 132 qwlcsdsklccpactfpgscclpcpgtcrpggygcgqegtrgsgndcqaigygea 191
 OY 177 CTDCMDGYFSSLRNETHSICTACDESCKTCSGLTNRDCGCEYGVWLDEGACVYDDECAA 236
 DB 192 cggcglygfeaeernashlvcsacfcgpcarcsqpeesncldckkqwalhnlkcvdidecgt 251
 OY 237 EPPPCSAOFCRKNANGSYTCEDSSCGCTGPGNCKECTCYAREHQCADVDECSL 296
 DB 252 egancgadqfcvntegsyecrdakacldcmgaqprckkcspgyqvgyskcldvdece- 310
 OY 297 AEKTCVRKNENCNTPGSYVCVPDGFEEEDACVP---PAEA---EATEGE 342
 DB 311 -tevcpgekqcenteggyrcicaegykkmeqglcvkeqldpesagffsemtede 362

RESULT 25

AAB80230
 ID AAB80230 standard; Protein; 420 AA.

XX
 AC AAB80230;

DT 24-APR-2001 (first entry)

XX Human PRO214 protein.

XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
 KW antiParkinsonian nootropic; neuroprotective; vulnerrary; cardiant;
 KW antiangiogenic; vasotrophic; antiasthmatic; antineumatic; cancer;
 KW antiarteritic; antiinfertility; antidiabetic; antifalral; diabetes;
 KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
 KW Ischaemia; Inflammation.

XX Homo sapiens.

PN WO200104311-A1.

PD 18-JAN-2001.

PF 22-FEB-2000; 2000WO-US04414.

XX
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 05-OCT-1999; 99WO-US21547.
 PR 29-NOV-1999; 99WO-US23089.
 PR 30-NOV-1999; 99WO-US28214.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 99WO-US30999.

PR 05-JAN-2000; 99WO-US00219.
 XX (GETH) GENENTECH INC.

XX Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Gdowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavlin IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WT.

DR WPI; 2001-081051/09.
 DR N-PSDB; AAF72391.

PT Sixty one nucleic acids encoding PRO polypeptides which are useful in
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
 PT Alzheimer's disease) -
 XX
 XX
 PS Claim 1; Fig 40; 39pp; English.

CC The present sequence is one of sixty one novel secreted and
 CC transmembrane PRO polypeptides. The PRO polypeptides are
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
 CC endometrial bleeding, angiodenesis, ischaemias such as coronary
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
 CC diabetes and retinal disorders such as retinitis pigmentosa.
 CC The PRO nucleic acids have applications in molecular biology, including
 CC use as hybridization probes, and in chromosome and gene mapping.

XX
 SQ Sequence 420 AA:

Query Match 48.5%; Score 971.5; DB 22; Length 420;
 Best Local Similarity 48.2%; Pred. No. 1.2e-60;
 Matches 170; Conservative 48; Mismatches 116; Indels 19; Gaps 5;

OY 7 AALGLPLLLLP-----PAPEAA--KKPPCHRCGLVDKFNQGVDTAKKNFGG 56
 DB 12 avlwgslflnlpplwlpqpsppppqphchrcglvdsfngkjlertlrdfnfgg 71
 OY 57 NTAMEKTLKYSSEIRLELEIGLCESSDFECNQMLEAQEHLAMWLQKSEYDLE 116
 DB 72 ntaweenlskydseclrvlelvgyvcsksdfechrlllelseelvswwlhkqgeapdlf 131
 OY 117 EMECVTLKVCSPGTYGPDCLACOGSGORPCSGNGHSGDGSROGSGSCRMHGYOGL 176
 DB 132 qwlcsdsklccpactfpgscclpcpgtcrpggygcgqegtrgsgndcqaigygea 191
 OY 177 EMECVTLKVCSPGTYGPDCLACOGSGORPCSGNGHSGDGSROGSGSCRMHGYOGL 236
 DB 192 cggcglygfeaeernashlvcsacfcgpcarcsqpeesncldckkqwalhnlkcvdidecgt 251
 OY 237 EPPPCSAOFCRKNANGSYTCEDSSCGCTGPGNCKECTCYAREHQCADVDECSL 296
 DB 252 egancgadqfcvntegsyecrdakacldcmgaqprckkcspgyqvgyskcldvdece- 310
 OY 297 AEKTCVRKNENCNTPGSYVCVPDGFEEEDACVP---PAEA---EATEGE 342
 DB 311 -tevcpgekqcenteggyrcicaegykkmeqglcvkeqldpesagffsemtede 362

RESULT 26

AAB48133
 ID AAB48133 standard; Protein; 420 AA.

XX AAB48133;

DT 02-APR-2001 (first entry)

XX

Human TANGO 206 variant 1 polypeptide.

TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse; transmembrane protein; antineuritic; cerebroprotective; arteriosclerosis; antiasthmatic; neuroprotective, cytosolic; cardiatic; hepatotropic; antileukemic; antidiabetic; antinfertility; antipyretic; vasotropic; antirheumatic; nephrotoxic; hemostatic; antitumor; osteopathic; ophthalmological; antischistosomal; antitumor; vulnerability; variant.

Homo sapiens.

Key Location/Qualifiers

Misc-difference 61 /label= E61D /note= "wild-type Glu is replaced by Asp"

WO200069885-A2.

23-NOV-2000.

15-MAY-2000; 2000WO-US13361.

14-MAY-1999; 99US-0312359.

(MILL-) MILLENNIUM PHARM INC.

Pan Y, Leiby KR; WPI: 2001-024999/03. N-PSDB; AAC84401.

Novel nucleic acids encoding secreted or transmembrane proteins, useful for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of the lung, liver, kidney or pancreas -

Claim 8; Page -: 209pp: English.

The invention provides human and mouse nucleic acids designated TANGO 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane proteins. The polypeptides, nucleic acids and their modulators may be useful for treating or modulating cholesterol uptake, blood coagulation, to modulate cell proliferation, morphogenesis and fate specification, tissue repair and renewal, to treat cancer and promote wound healing, modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan syndrome, protein S deficiency, modulate allergic or inflammatory response, acid secretion, tropic effects on gastrointestinal mucosa, and promote ulcer healing, treat bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis, leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain herniations, meningitis, ischemic brain or heart disease, infection, intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart disease, pulmonary heart disease, rheumatic fever, congenital heart disease, myocardial disease, atherosclerosis, hypertension, jaundice, hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome, sickle cell disease, renal failure, ischemic bowel disease, Crohn's disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome, neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune Albright syndrome, infertility, uterine disorders, viral disease. The present sequence represents a human TANGO 206 variant polypeptide. Note: the present variant sequence has been constructed using the information provided in the specification.

	Query Match	48.5%	Score 971.5;	DB 22;	Length 420;
	Best Local Similarity	48.2%;	Pred. No. 1.2e-60;		
	Matches 170; Conservative	48;	Mismatches 116;	Indels 19;	Gaps 5;
Oy	7 AALGLPLLLLP-----PAPEAA-KRPTPCRGRGLVDKFNQGVADTKAKNFGGG	56			
	: : : : :				
Db	12 avlwgisiflnlpgrpwqpsppgsppsgphpcctcrglvdstinkgldrtldrfnfggg	71			

OY 57 NTAMEKRLSKYSESEIRLLEIIELGLDESSPFECNOMLAEQHEHLEAMWLQQLKSEYPDLF 116
 Db 72 nlaveeenlslkykdssetrlveelvgvsksdfechrlllseeliveswwfhkqgeapdlf 131
 OY 117 EMFVCYKTLKVCSCSGTGYPPDCLACQGSORPCSCNGHSCSDSGROGSGSCRMHGYOGPL 176
 Db 132 gwlcscdsklkcpagtlfpqscplpcpggterpcpgygqcegegrtgrsgshcdqagaygaa 191
 OY 177 CTDGMDGYFSSLRNETHSICTACDESECKTCSGLTNRDGCECEVGGWLDEGACVYDDECAA 236
 Db 192 cggcglygfseearnashlvcasacfpqarcsqpsseencldqckkqwalhhlkcvldieqg 251
 OY 237 EPPPCSAQPFCKNANGSTCEHCSSCVCGCTGEBPQNCRCICISYAREHQCADVDVDCSL 296
 Db 232 egangagdqgfvntegyspccrcaakalqumgagpgrckkcspgyqvgyskcldvdece- 310
 OY 297 AEKTCVRNENCMYNTPGSYVCVCPDGEFEEDACVP---PAEA----EATEGE 342
 Db 311 -tevcpgeknkgcenteggyrticicaeygkqmeglcvkqepesdgffsemtede 362
 RESULT 27
 AAB27228
 ID AAB27228 standard; Protein; 420 AA.
 XX AAB27228;
 AC
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human EXMAD-6 SEQ ID NO: 6.
 XX
 KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 inflammation; reproductive disorder; cardiovascular disorder;
 immune disorder; musculoskeletal disorder; developmental disorder;
 gastrointestinal disorder; cell proliferation disorder.
 XX
 OS Homo sapiens.
 OS
 PN WO200068380-A2.
 PN
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000WO-US12811.
 XX
 PR 11-MAY-1999; 99US-0133643.
 PR 23-AUG-1999; 99US-0150409.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DM;
 PI Azimzai Y;
 XX
 DR WPI: 2001-007395/01.
 DR N-PSDB: AAC66895.
 XX
 PT Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX
 Claim 1; Page 93-94; 129pp; English.
 XX
 CC The present invention provides the protein and coding sequences for 25
 CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.
 XX
 Sequence 420 AA;
 50

CC The present invention provides the protein and coding sequences for 25
CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC useful in the prevention and treatment of cancers, cell proliferation,
CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC gastrointestinal disorders and inflammation.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS
XX Example 2; SEQ ID NO 6616; 10078bp; English.
CC
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nocitropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX
SQ Sequence 513 AA;

Dd	105	aalwlsflfllnlpplwlqppppgaspqpqhpcntcrglydsfnkglertlrdfngg	164
Oy	57	NPAWEKTLTKESSEIRLEILEIGCESSDPECNOMLEAEOEHLEAWMLQLKSEYPLDF	116
Dd	169	nlaweenalskykdssetrlvevlegvcasdfsctehnllelseelweswfhnkgqeapdlf	224
Oy	117	EWFECYKTLKVCSSPSTGYGPDCLACOGGSORPCPSGNHCGSDSRGDDSGCRCHMGYQGPL	176
Dd	225	gylcdslstlclcpagultfpsscldpcpgyttertpcggyqgcgeglrtgsgndcdqa9y9ga	284
Oy	177	CYDCKMDGYFSSLRNETHSICTPACDESCSKCSGLTNRDCCECEGWLDBAGCAVDVECAA	236
Dd	285	cggcglglyleaernashlvcsacifpccatcspgspeenclcqckgwahlhkcvdidcegt	344
Oy	237	EPPPCSAAOFCCNANGSYICECDSSCVCTGBEGPNCKECTISGVAREHGQCADVDEGL	296
Dd	345	eganagadqlfcyntlegsyecrdcaaclqcmagapbrckcpspyqvyqyskctldvece-	403
Oy	297	AERTVKRRKENNYNTPGSYVYCVPDOFEETEDACVP---PAEA---EATGEF	342
Dd	404	-tevpgenkqgentcegryrcicaegykmeglcvkqglpeaaagffisemtede	455
 RESULT 29 AAB48134			
ID	AAB48134	standard; Protein: 420 AA.	
XX AC	AAB48134;		
XX DT	02-APR-2001	(first entry)	
XX DE	Human TANGO 206 variant 2 polypeptide.		
KW KW	TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;		
KW KW	transmembrane protein; antiinemic; cerebroprotective; arteriosclerosis;		
KW KW	antiasthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;		
KW KW	antinflammatory; antididiabetic; antifertility; antipyletic; vasotrophic		
KW KW	antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathnic;		
XX XX	ophthalmological; antisickling; antilucer; vulnerary; variant.		
OS	Homo sapiens.		

```

XX Key Location/Qualifiers
FH Misc-difference 76 /label= E76D
FT /note= "wild-type Glu is replaced by Asp"
XX
XX WO200069885-A2.
XX
XX 23-NOV-2000.
XX
XX 15-MAY-2000; 2000MO-US13361.
XX
XX 14-MAY-1999; 99US-0312359.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lei by KR;
XX
XX WPI: 2001-024999/03.
XX
XX N-PSDB: AAC84402.
XX
XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
XX the lung, liver, kidney or pancreas -
XX
XX Claim 8; Page -: 209pp; English.
XX
XX The invention provides human and mouse nucleic acids designated TANGO
XX 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
XX proteins. The polypeptides, nucleic acids and their modulators may be
XX useful for treating or modulating cholesterol uptake, blood coagulation,
XX to modulate cell proliferation, morphogenesis and fate specification,
XX tissue repair and renewal, to treat cancer and promote wound healing,
XX modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
XX syndrome, protein S deficiency, modulate allergic or inflammatory
XX response, acid secretion, tropic effects on gastrointestinal mucosa, and
XX promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
XX fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
XX leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
XX herniations, meningitis, ischemic brain or heart disease, infarction,
XX intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
XX disease, pulmonary heart disease, rheumatic fever, congenital heart
XX disease, myocardial disease, atherosclerosis, hypertension, jaundice,
XX hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
XX sickle cell disease, renal failure, ischemic bowel disease, Crohn's
XX disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
XX neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
XX Albright syndrome, infertility, uterine disorders, viral disease. The
XX present sequence represents a human TANGO 206 variant polypeptide.
XX Note: the present variant sequence has been constructed using the
XX information provided in the specification.
XX
XX Sequence 420 AA:

```

```

Query Match 48.4%; Score 969.5; DB 22; Length 420;
Best Local Similarity 47.9%; Pred. No. 1,6e-60;
Matches 169; Conservative 49; Mismatches 116; Indels 19; Gaps 5;

OY 7 AATGTLPLLLLP-----PAPAA--KKPTCHRCRLVKNFNGMDTAKKNGG 56
DB 12 avtawgslfthlpqplwlgppppspgppchrcrjvdsfnqglertlrtdnpgg 71
OY 57 NTFMWEKTLKYSSEIRLEIEGLCESSDFPCNQMLNQEHLLEAMWLQKSEYDFL 116
DB 72 ntawdeenlskydsetrlvelegvsksdfechrllseelvevwvfhkqgepdltf 131
OY 117 EMFVCVTKLVKCSBPYTGPDCLACQGGSGRCGNGHCSGSDSGRQDGSRCRHMGYQGPL 176
DB 132 gvlcsdsklkcpcagtpgpcclpcpggterpcggygqcegegrtrgsghdcqagyygea 191
OY 177 CTGCMGYSFSLRNETHSTACDESKTCSGLNRDGCCEGVWVLDACADVDECA 236
DB 177 CTGCMGYSFSLRNETHSTACDESKTCSGLNRDGCCEGVWVLDACADVDECA 236

```

```

DB 192 cggcg1gyfeaeernaahlycsacifgpacrcsgpeesnc1qckkya1hllkcvdidegt 251
OY 237 EPPPCSAAGFCCKNANGSYTCCECDSSCVGCTGEGPNCCKECISGYAREHGQCADVDECSL 296
DB 252 egancgqadqfcvntegsyecrcdcaclcmagpprcckcspgyqgyskclddece- 310
OY 297 AEKTCVRKNKNCYNTPGYSVVCVCPDGFEEEDACVP---PAAE---EATGE 342
DB 311 -tevcpngkqcenteggyrcloaegykqmegicvkeqjpesagffsemtede 362

RESULT 30
AAB48135
ID AAB48135 standard; Protein; 420 AA.
XX
AC AAB48135;
XX
DT 02-APR-2001 (first entry)
XX
XX Human TANGO 206 variant 3 polypeptide.
XX
XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
XX transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
XX antiasthmatic; neuroprotective; cytosstatic; cardiast; hepatotropic;
XX antiinflammatory; antidiabetic; antiinfertility; antipyletic; vasotropic;
XX antirheumatic; nephrotropic; hemostatic; antilipemic; osteophtic;
XX ophthalmological; antislacking; antilucer; vulnerary; variant.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 77 /label= E77D
FT /note= "wild-type Glu is replaced by Asp"
XX
XX WO200069885-A2.
XX
XX 23-NOV-2000.
XX
XX 15-MAY-2000; 2000MO-US13361.
XX
XX 14-MAY-1999; 99US-0312359.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lei by KR;
XX
XX WPI: 2001-024999/03.
XX
XX N-PSDB: AAC84403.
XX
XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
XX the lung, liver, kidney or pancreas -
XX
XX Claim 8; Page -: 209pp; English.
XX
XX The invention provides human and mouse nucleic acids designated TANGO
XX 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
XX proteins. The polypeptides, nucleic acids and their modulators may be
XX useful for treating or modulating cholesterol uptake, blood coagulation,
XX to modulate cell proliferation, morphogenesis and fate specification,
XX tissue repair and renewal, to treat cancer and promote wound healing,
XX modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
XX syndrome, protein S deficiency, modulate allergic or inflammatory
XX response, acid secretion, tropic effects on gastrointestinal mucosa, and
XX promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
XX fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
XX leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
XX herniations, meningitis, ischemic brain or heart disease, infarction,
XX intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
XX disease, pulmonary heart disease, rheumatic fever, congenital heart
XX disease, myocardial disease, atherosclerosis, hypertension, jaundice,
XX hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,

```

CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
 CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
 CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
 CC Albright syndrome, infertility, uterine disorders, viral disease. The
 CC present sequence represents a human TANGO 206 variant polypeptide.
 CC Note: the present variant sequence has been constructed using the
 CC information provided in the specification.

XX Sequence 420 AA;

Query Match

Best Local Similarity 48.4%; Score 969.5; DB 22; Length 420;
 Pred. No. 1.6e-60;
 Matches 169; Conservative 49; Mismatches 116; Indels 19; Gaps 5;

QY 7 AALGLPLULLLP-----PAPEAA--KKPTCHRCRLVDFKNGAMDYAKKNGCG 56
 DB 12 avlwgislflnlpqplwlpqpppsppqphchrcrglvdslfnglertlrdnfg99 71
 QY 57 NTAMEKTLKSYESSSEIRLEILEGLCESSDFECNQMLEAOEHLLEAWMLQKSEYDLE 116
 DB 72 ntawedenlskykdserrlvelegvcsksdfchrlleseeelveswvfkkgaeapdlf 131
 QY 117 EMFCVKTLYKVCSPGTYGPDCLACQGSQRPSCNGHCSGDSGSGRCHMGYOGPL 176
 DB 132 qwlcsdsklkcpagltfpgscldpqpgrterpcggygqcegegrtrgsghdcqagysgea 191
 QY 177 CTQCMGYSFSLRNETHSICACDESKTSGLTNRDCEGEVGVWVDEGACVDDECAA 236
 DB 192 cggcgaglyfleaernashlvcsacifgpcarcspeesncldckkqywalhhlkcvdidecgt 251
 QY 237 EPPPSAAGFCFKNANGSYTCCECDSSCVGCTGEGPGNCKECISGYAREHGOACADVDCSL 296
 DB 252 egangcgadqfvcntegyecrdcaekalgcmagaprcckkcpgyqgyskclvddece- 310
 QY 297 AEKTCVRKNENCYNTPGSYVCPCPDGFETEDACVP--PAEA---EATGE 342
 DB 311 -tevcpgekqcenteggyrcicaeykkqmeglcvckeqipesagffseamlede 362

RESULT 31

AAB38395
 ID AAB38395 standard; Protein: 434 AA.

XX AAB38395;

DT 31-JAN-2001 (first entry)

DE Human secreted protein encoded by gene 56 clone HDTAT90.

XX Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
 KM cytostatic; cardiant; vasotrophic; cerebroprotective; neuroprotective;
 KM neotrophic; antibacterial; vitucide; fungicide; optalmalogical; human;
 KM vulnerary; gene therapy; infection; secreted protein.

OS Homo sapiens.

PN WO2000061623-A1.

PD 19-OCT-2000.

PF 06-APR-2000; 2000WO-US08979.

PR 09-APR-1999; 99US-0128693.

PR 26-APR-1999; 99US-0130991.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Ni J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;

PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;
 PI Young PE;
 DR WPI; 2000-647418/62.

XX New nucleic acid molecules encoding 62 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PS Claim 11; Page 652-653; 716pp; English.

XX Sequences AAB38321-B38396 represent the amino acid sequences of 62

CC human secreted proteins encoded by the genes AAC69512-C69587. The genes
 CC and proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from
 CC a range of human tissues disclosed in the specification. The nucleic
 CC acids, proteins, antibodies and (ant)agonists are useful in the
 CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
 CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms
 CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac
 CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
 CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
 CC infections caused by bacteria, viruses and fungi; and (h) ocular
 CC disorders e.g. corneal infection. The polypeptides can also be used to
 CC aid wound healing and epithelial cell proliferation, to prevent skin
 CC aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis.

SQ Sequence 434 AA;

Query Match

Best Local Similarity 48.3%; Score 969; DB 21; Length 434;
 Pred. No. 1.6e-60;
 Matches 168; Conservative 50; Mismatches 119; Indels 20; Gaps 4;

QY 7 AALGLPLULLLP-----PAPEAA--KKPTCHRCRLVDFKNGAMDYAKKNGCG 56
 DB 9 avlwgislflnlpqplwlpqpppsppqphchrcrglvdslfnglertlrdnfg99 68
 QY 57 NTAMEKTLKSYESSSEIRLEILEGLCESSDFECNQMLEAOEHLLEAWMLQKSEYDLE 116
 DB 69 ntawedenlskykdserrlvelegvcsksdfchrlleseeelveswvfkkgaeapdlf 128
 QY 117 EMFCVKTLYKVCSPGTYGPDCLACQGSQRPSCNGHCSGDSGSGRCHMGYOGPL 176
 DB 129 qwlcsdsklkcpagltfpgscldpqpgrterpcggygqcegegrtrgsghdcqagysgea 188
 QY 177 CTQCMGYSFSLRNETHSICACDESKTSGLTNRDCEGEVGVWVDEGACVDDECAA 236
 DB 189 cggcgaglyfleaernashlvcsacifgpcarcspeesncldckkqywalhhlkcvdidecgt 248
 QY 237 EPPPSAAGFCFKNANGSYTCCECDSSCVGCTGEGPGNCKECISGYAREHGOACADVDCSL 296
 DB 249 egangcgadqfvcntegyecrdcaekalgcmagaprcckkcpgyqgyskclvddece- 307
 QY 297 AEKTCVRKNENCYNTPGSYVCPCPDGFETEDACVPAAEATGEGSPQTOLPSREDL 353
 DB 308 -tevcpgekqcenteggyrcicaeykkqmeglc-----kqjpagtfltdl 355

RESULT 32

AAY76151
 ID AAY76151 standard; Protein: 434 AA.

XX AAY76151;

DT 23-MAR-2000 (first entry)

DE Human secreted protein encoded by gene 28.

XX Human; secreted protein; cancer; tumour; developmental abnormality;
 KM foetal deficiency; blood disorder; immune system disorder; inflammation;
 KM autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KM schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KM atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KM digestive disorder; endocrine disorder; infection; AIDS; leukaemia;

CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC Albright syndrome, infertility, uterine disorders, viral disease. The
CC present sequence represents the mouse TANGO 206 polypeptide.

XX Sequence 420 AA;

Query Match

Best Local Similarity 48.3%; Score 968.5; DB 22; Length 420;
Pred. No. 1.9e-60;
Matches 171; Conservative 43; Mismatches 121; Indels 29; Gaps 5;

QY 3 LPRRAALGLPLULLL-----PPAPAAKPPCHRCGLVKNQGM 45
DB 4 lprp---glvpsllwclslflslpgpwwlqpspphpspraephchrcrlvlnfnkyl 60
QY 46 VDTAKKNGGNTAMEEKTLSKYESSERILLEILEGLCESSDFPCNOMLEAOEHLLEAMW 105
DB 61 ertlrndfgygnntaweeklskydsetrlvelegvcsrdsfchrllelseelvenw 120
QY 106 LQLKSEYVDLFEMFCVTKLKYCSPGTYPDCLACOGSGORPCSGNGHCSGDSGNOGDS 165
DB 121 fhrgqeadlflqwlcsdsklkcpsgtfpgscplpcpgtcrpccggygcgegrtgsqh 180
QY 166 CRCHMGYOGPLCTDCMNGYRSSLNETHSITACDESCRTSGLTNRDCCGEVGMWIDE 225
DB 181 cdcqagyggecgcgylgfaeernshlvcsacfgpcarctgpeeshclqckkgywalhh 240
QY 226 GACVDVDECAAEPPPCSAOFCCKNANGSYNCECDSSCVGCTGSGPGNCKECISGYAREH 285
DB 241 lkcvdidecgtcgatcgadqfcvntleesyeccrdakacolgmgagpgrckkcsrgyqvg 300
QY 286 GQCADVDECSLAETCYARKNENCYNTPGSYVCVPDGFETEDAC---VPPAE---AEA 338
DB 301 skcidvdecetv--vcpgenekcenteggyrcvcagayrgedglcvkqypesagffaem 358
QY 339 TEGE 342
DB 359 TEGE 362

RESULT 34

AAB48137
ID AAB48137 standard; Protein: 420 AA.

XX AAB48137;

DT 02-APR-2001 (first entry)

DE Mouse TANGO 206 variant 2 polypeptide.

XX TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
XX transmembrane protein; anilemic; cerebroprotective; arteriosclerosis;
XX antisthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;
XX antiinflammatory; antidiabetic; antifertility; antipyletic; vasotropic;
XX antineumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
XX ophthalmological; antisticking; antulcer; vulnerary; variant.

OS Mus sp.

XX Key Location/Qualifiers

FT Misc-difference 61

FT /label= E61D

XX /note= "wild-type glu is replaced by asp"

PN MO200069885-A2.

PD 23-NOV-2000.

XX 15-MAY-2000; 2000MO-US13361.

XX 14-MAY-1999; 99US-0312359.
XX (MILL-) MILLENNIUM PHARM INC.
XX Pan Y, Leiby KR;

XX WPI: 2001-024999/03.

XX N-PSDB: AAC84405.

XX Novel nucleic acids encoding secreted or transmembrane proteins, useful
XX for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
XX the lung, liver, kidney or pancreas -

PS Claim 8; Page -: 209pp; English.

CC The invention provides human and mouse nucleic acids designated TANGO
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
CC proteins. The polypeptides, nucleic acids and their modulators may be
CC useful for treating or modulating cholesterol uptake, blood coagulation,
CC to modulate cell proliferation, morphogenesis and fate specification,
CC tissue repair and renewal, to treat cancer and promote wound healing,
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
CC syndrome, protein S deficiency, modulate allergic or inflammatory
CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC herniations, meningitis, ischemic brain or heart disease, infarction,
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC disease, myocardial heart disease, rheumatic fever, congenital heart
CC disease, pulmonary disease, atherosclerosis, hypertension, jaundice,
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC disease, hernias, hyperadrenalism, hyperadrenalism, Cushing's syndrome,
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC Albright syndrome, infertility, uterine disorders, viral disease. The
CC present sequence represents a mouse TANGO 206 variant polypeptide.
CC Note: the present variant sequence has been constructed using the
CC information provided in the specification.

XX Sequence 420 AA;

Query Match

Best Local Similarity 48.3%; Score 967.5; DB 22; Length 420;
Pred. No. 2.3e-60;
Matches 171; Conservative 43; Mismatches 121; Indels 29; Gaps 5;

QY 3 LPRRAALGLPLULLL-----PPAPAAKPPCHRCGLVKNQGM 45
DB 4 lprp---glvpsllwclslflslpgpwwlqpspphpspraephchrcrlvlnfnkyl 60
QY 46 VDTAKKNGGNTAMEEKTLSKYESSERILLEILEGLCESSDFPCNOMLEAOEHLLEAMW 105
DB 61 ertlrndfgygnntaweeklskydsetrlvelegvcsrdsfchrllelseelvenw 120
QY 106 LQLKSEYVDLFEMFCVTKLKYCSPGTYPDCLACOGSGORPCSGNGHCSGDSGNOGDS 165
DB 121 fhrgqeadlflqwlcsdsklkcpsgtfpgscplpcpgtcrpccggygcgegrtgsqh 180
QY 166 CRCHMGYOGPLCTDCMNDYFSSLNETHSITACDESCRTSGLTNRDCCGEVGMWIDE 225
DB 181 cdcqagyggecgcgylgfaeernshlvcsacfgpcarctgpeeshclqckkgywalhh 240
QY 226 GACVDVDECAAEPPPCSAOFCCKNANGSYNCECDSSCVGCTGSGPGNCKECISGYAREH 285
DB 241 lkcvdidecgtcgatcgadqfcvntleesyeccrdakacolgmgagpgrckkcsrgyqvg 300
QY 286 GQCADVDECSLAETCYARKNENCYNTPGSYVCVPDGFETEDAC---VPPAE---AEA 338
DB 301 skcidvdecetv--vcpgenekcenteggyrcvcagayrgedglcvkqypesagffaem 358
QY 339 TEGE 342

```

DB      359 tede 362

RESULT 35
AAB48136
ID      AAB48136 standard; Protein; 420 AA.
XX
AC      AAB48136;
XX
DT      02-APR-2001 (first entry)
XX
DE      Mouse TANGO 206 variant 1 polypeptide.
XX
KW      TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW      transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
KW      antiasthmatic; neuroprotective; cytosatic; cardiatic; hepatotropic;
KW      antiinflammatory; antidiabetic; antifertility; antipyretic; vasotropic;
KW      antirheumatic; nephrotropic; hemostatic; antileptic; osteopathic;
KW      ophthalmological; antisickling; antilucer; vulnerary; variant.
XX
OS      Mus sp.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 42 /label= E42D
FT      /note= "Wild-type Glu is replaced by Asp"
XX
PN      WO200069885-A2.
XX
PD      23-NOV-2000.
XX
PF      15-MAY-2000; 2000WO-US13361.
XX
PR      14-MAY-1999; 99US-0312359.
XX
PA      (MILL-) MILLENNIUM PHARM INC.
XX
PI      Pan Y, Leiby KR;
XX
DR      WPI: 2001-024999/03.
DR      N-PSDB; AAC84404.
XX
PT      Novel nucleic acids encoding secreted or transmembrane proteins, useful
PT      for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
PT      the lung, liver, kidney or pancreas -
XX
PS      Claim 8; Page -; 209pp; English.
XX
XX      The invention provides human and mouse nucleic acids designated TANGO
XX      204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
XX      proteins. The polypeptides, nucleic acids and their modulators may be
XX      useful for treating or modulating cholesterol uptake, blood coagulation,
XX      to modulate cell proliferation, morphogenesis and fate specification,
XX      tissue repair and renewal, to treat cancer and promote wound healing,
XX      modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
XX      syndrome, protein S deficiency, modulate allergic or inflammatory
XX      response, acid secretion, tropic effects on gastrointestinal mucosa, and
XX      promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
XX      fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
XX      leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
XX      herniations, meningitis, ischemic brain or heart disease, infarction,
XX      intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
XX      disease, pulmonary heart disease, rheumatic fever, congenital heart
XX      disease, myocardial disease, atherosclerosis, hypertension, jaundice,
XX      hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
XX      sickle cell disease, renal failure, ischemic bowel disease, Crohn's
XX      disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
XX      neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
XX      Albright syndrome, infertility, uterine disorders, viral disease. The
XX      present sequence represents a mouse TANGO 206 variant polypeptide.
XX      Note: the present variant sequence has been constructed using the
XX      information provided in the specification.

```

```

XX      SQ      Sequence      420 AA;
XX
XX      Query Match      48.2%; Score 966.5; DB 22; Length 420;
XX      Best Local Similarity 47.0%; Pred. No. 2.7e-60;
XX      Matches 171; Conservative 42; Mismatches 122; Indels 29; Gaps 5;
XX
XX      3 LPRRALGLPLDLLL-----PPAPEAAKKPTPCRCRGIVAKFNQGM 45
XX      4 Lprr---glvpsllwclsiflslppvwlqpppphpsrpadphchtrralvdnfnkgl 60
XX
XX      46 VDTAKNFGGGTAMPEKTLSTKESSEIRLLLEBGLCSSDFECNQMTBADEHLEAWM 105
XX      61 erlirnfgygnaweeeklskykdsrtrlvlevlegvcsrdsfchrlilleseelvenw 120
XX
XX      106 LQKSEYPLFEMFCKTLKVCSSPGTYGPDCLACGGSGORPCSGHSGDSRGDGS 165
XX      121 fhrgqapdlfgwlcadsxlkccpsglfpgscilpcpggterpcgygqcegeglr99sgh 180
XX
XX      166 CRCHMGYGPFLCTDCMDGFFSSLRNETHSICTACDSCSKTCSGLTRNDCGEGEVGWLDE 225
XX      181 cdqgagyggeacgqcglylfeaeernshlvcaacifgparctgpeeschlgckkqyalnh 240
XX
XX      226 GACVDVDECAAEPPPCSAOFCNANGSTYCEBDSSCYVGTGEGGNCKECTSGTAREH 285
XX      241 lkcvdidecgeqatgcagdqfcvntegsyecrdcakacigcmgaspgrckkcsrygqvg 300
XX
XX      286 GGCADVDECSLAKTGVNRKNENCYNTPGSYVCVCPGFEETEDAC---VPPE---AEA 338
XX      301 skcldvdececv--vcpgenekcenteggyrcvcaegyrgedlcvkeqypesaagffaeM 358
XX
XX      QY      339 TEGE 342
XX      DB      359 tede 362
XX
XX      RESULT 36
XX      AAB48138
XX      ID      AAB48138 standard; Protein; 420 AA.
XX
XX      AC      AAB48138;
XX
XX      DT      02-APR-2001 (first entry)
XX
XX      DE      Mouse TANGO 206 variant 3 polypeptide.
XX
XX      KW      TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
XX      transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
XX      antiasthmatic; neuroprotective; cytosatic; cardiatic; hepatotropic;
XX      antiinflammatory; antidiabetic; antifertility; antipyretic; vasotropic;
XX      antirheumatic; nephrotropic; hemostatic; antileptic; osteopathic;
XX      ophthalmological; antisickling; antilucer; vulnerary; variant.
XX
XX      OS      Mus sp.
XX
XX      FH      Key Location/Qualifiers
XX      FT      Misc-difference 76 /label= E76D
XX      FT      /note= "Wild-type Glu is replaced by Asp"
XX
XX      PN      WO200069885-A2.
XX
XX      PD      23-NOV-2000.
XX
XX      PF      15-MAY-2000; 2000WO-US13361.
XX
XX      PR      14-MAY-1999; 99US-0312359.
XX
XX      PA      (MILL-) MILLENNIUM PHARM INC.
XX
XX      PI      Pan Y, Leiby KR;
XX

```

DR WPI: 2001-024999/03.
DR N-PSDB: AAC84406.
XX
PT Novel nucleic acids encoding secreted or transmembrane proteins, useful
PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
PT the lung, liver, kidney or pancreas -
XX
PS Claim 8; Page 7; 209pp; English.

XX The invention provides human and mouse nucleic acids designated TANGO
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
CC proteins. The polypeptides, nucleic acids and their modulators may be
CC useful for treating or modulating cholesterol uptake, blood coagulation,
CC to modulate cell proliferation, morphogenesis and fate specification,
CC tissue repair and renewal, to treat cancer and promote wound healing,
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
CC syndrome, protein S deficiency, modulate allergic or inflammatory
CC response, acid secretion, trophic effects on gastrointestinal mucosa, and
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC herniations, meningitis, ischemic brain or heart disease, infarction,
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC disease, pulmonary heart disease, rheumatic fever, congenital heart
CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC Albright syndrome, infertility, uterine disorders, viral disease. The
CC present sequence represents a mouse TANGO 206 variant polypeptide.
CC Note: the present variant sequence has been constructed using the
CC information provided in the specification.
XX
SQ Sequence 420 AA;

Query Match 48.2%; Score 965.5; DB 22; Length 420;
Best Local Similarity 46.7%; Pred. No. 3.1e-60;
Matches 170; Conservative 44; Mismatches 121; Indels 29; Gaps 5;

QY 3 LPRRAAGLPLLL-----PPAPAAKKPPPHRCRGIVDKFNQGM 45
DB 4 LPPR---GLVPSLLWCISLFLSLPGPWLQPSPPHPSPREHPCHCRALVDNFINKI 60
QY 46 VDTAKKNGGNTAMEKTLKSYSESEIRLEILEGICSSDFECNOMLEAOEHLZAMW 105
DB 61 ertirnfgygnltawdeeklskydserrlvelegvcsrtdfechrllseelvenwv 120
QY 106 LQAKSEYPDLEFEMPCVTKLKYCCSPGYTPDCLACQGSQRPSCGNGHCSGDSHQDGS 165
DB 121 fhqgseapdlftqwlscsdsklkcpsgtfpcplpcpggtterpcggygcegegrtgsgh 180
QY 166 CRCHMGVQGLPCTDCMDCGYFSSLNETHSTCTACDESKTSCGLTNRCGECENGVWUDE 225
DB 181 cdcaagyyggaagcgqjlyfleaernshlvcsactfgyrcartcgypeesnlckkqkwalh 240
QY 226 GACVDVDECAEPPTCAAFCKANAGSYTCEDDSCVGTGEGSPGCKICISYAREH 285
DB 241 lkcvdidecgeatcgadqfclvntegseyecrdckacilgcmagpgrckcsryqgvg 300
QY 286 GQCADVDECSLAERTCVKRNENCYNTPGSYVVCVCPDGFEEEDAC---VPPAE---AEA 338
DB 301 skclvdceetv---vcpgekenkcenteggyrcvcaegyrtdgdglcvkqepesagffiem 358
QY 339 TEGE 342
DB 359 tede 362

RESULT 37
AAB38394
ID AAB38394 standard; Protein: 392 AA.

XX
AC AAB38394;
XX
DT 31-JAN-2001 (first entry)
XX
DE Human secreted protein encoded by gene 56 clone HPRAL78.
XX
DE
XX
KW Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;
KW nootropic; antibacterial; virucide; fungicide; ophthalmological; human;
KW vulnertary; gene therapy; infection; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200061623-A1.
XX
PD 19-OCT-2000.
XX
PE 06-APR-2000; 2000WO-US08979.
XX
PR 09-APR-1999; 99US-0128693.
PR 26-APR-1999; 99US-0130991.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, N1 J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;
PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;
PI Young PE;
XX
DR WPI: 2000-647418/62.

PT New nucleic acid molecules encoding 62 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS
XX
SQ Claim 11; Page 650-651; 716pp; English.

XX Sequences AAB38321-B38396 represent the amino acid sequences of 62
CC human secreted proteins encoded by the genes AAC69512-C69587. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms
CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac
CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
CC infections caused by bacteria, viruses and fungi; and (h) ocular
CC disorders e.g. corneal infection. The polypeptides can also be used to
CC aid wound healing and epithelial cell proliferation, to prevent skin
CC aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis.
XX
SQ Sequence 392 AA;

Query Match 42.9%; Score 860.5; DB 21; Length 392;
Best Local Similarity 44.2%; Pred. No. 7.1e-53;
Matches 156; Conservative 45; Mismatches 105; Indels 47; Gaps 6;

QY 7 AALGLPLPLLLP-----PAPRA--KKPTPHRCRGIVDKFNQGMVDTAKKNGFGG 56
DB 12 avlwglsiflnlppplwlgpspppspppqpnphtctrglydsfnkjetrlttdn 71
QY 57 NTAMEKTLKSYSESEIRLEILEGICSSDFECNOMLEAOEHLZAMWLQAKSEYPDLE 116
DB 72 ntaweenlskydserrlvelegvcskdsdfechrllleaelveswfwfhkgqapdlf 131
QY 117 EMFCVTKLKYCCSPGYTPDCLACQGSQRPSCGNGHCSGDSRCRCHMGVQGLP 176
DB 132 qwlcsdsklkcpsgtfpcplpcpggtterpcggygcegegrtgsghcdcaagygaa 191


```

Db 143 lldsltpshlpkwhespcgtpeapgnslgflckfnfkymcrlalg-gpgrvtyt 201
QY 62 ---EKTLSKSESEIRLLEILEGLC-----ESSDFECNOMLEAQBHLBAMWLQKSEY 112
Db 202 tpfqatsslea--vpfasvanvacgdeaksethylene-----kt 241
QY 113 PDLEFMPCKYKTLKVCSPGTYGPDCLACGGSGRPCS-GNGHSGSGSROGDGS--CRCH 169
Db 242 pglfhw-----gssgplcvspkfg-----csfnnggcqgdcfegdgdsfrgcgr 285
QY 170 MGYO-----GPLCTD-----CMDGYFSSLRNETHSI--- 195
Db 286 pglrllldlvtcasrnpssnptctggmchsvplsenytrcpsyql-d-sgvyhcvdld 344
QY 196 -C--TACDESKCTKSGLTNRDCECEVGVWLD--EGACVDVDECAEPPPCSAOFCNK 249
Db 345 ecqdspcaqdcvntlg--sfhc-ecwvygypspkceacedvdecaaanpc--agcgin 399
QY 250 ANGSTYCECDSSCGCTGEGPNCCKECLISGY--AREHGQCADVDECSLAE-KTCVRKN 305
Db 400 tdgsfyc-----scke-----gyivsgedstqcedidcedsdargnpc---d 437
QY 306 ENCYNTPGSVVCVCPDGFETEDACVPAEAETEG---ESPQLPSRED 352
Db 438 slcfnldgsfrgcgppgwe-----lapngvfcsrtglvsefpaprpqked 482

RESULT 45
AA79193
ID AA79193 standard; protein; 644 AA.
XX
AC AA79193;
XX
DT 19-JUN-2000 (first entry)
XX
DE Haematopoietic stem cell specific protein AA4.
XX
KW AA4; haematopoietic stem cell; immune system disorder;
KM leukaemia; anti-leukaemic; immunomodulator; therapy; mouse.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Reptide 1..19 /note="signal peptide"
FT Protein 20..644 /note="mature protein"
FT Domain 35..129 /note="C-type lectin carbohydrate recognition d"
FT Domain 575..597 /note="transmembrane domain"
FT Region 282..297 /note="EGF-like repeat"
FT Region 325..340 /note="EGF-like repeat"
FT Region 451..464 /note="EGF-like repeat"
FT Region 343..364 /note="EGF-like repeat"
FT Region 382..406 /note="calcium-binding EGF motif"
FT Region 424..449 /note="calcium-binding EGF motif"
FT Region /note="calcium-binding EGF motif"
XX
PN MO200011168-A2.
XX
PD 02-MAR-2000.
XX
PF 20-AUG-1999; 99WO-US19052.
XX
PR 21-AUG-1998; 98US-0138132.
XX
PA (UYPR-) UNIV PRINCETON.

```

```

XX
PI Lemischka I, Moore K;
XX
DR WPI: 2000-237650/20.
XX
DR N-PSDB; AA294131.
XX
PT Hematopoietic stem cell signaling proteins modulating replication and
XX differentiation for treating immune system disorders and leukaemia -
XX Claim 21; Fig 9; 256pp; English.
XX
CC The present sequence is that of mouse haematopoietic stem cell
CC (HSC) specific protein AA4, a type I transmembrane protein that
CC shows homology to C1qR, the human receptor for complement C1q.
CC AA4 is a preferred example of claimed HSC-specific proteins (see
CC AA79193-93) predicted from novel isolated HSC-specific nucleic acids
CC (see AA294077-131). The HSCs are especially primitive HSCs (PHSCs)
CC such as umbilical cord cells, bone marrow cells and foetal liver
CC cells. The encoded proteins are growth factors, transcription
CC factors, splicing factors, capping factors, transport proteins,
CC translation factors or replication factors that modulate HSC
CC activity, especially differentiation or replication. The invention
CC provides claimed methods for identifying PHSC-specific nucleic
CC acids; for generating a stem cell/progenitor cell from PHSCs; for
CC identifying the presence of a PHSC in a sample; for identifying the
CC presence in a sample of a compound that modulates HSC activity; for
CC using such a compound to treat an immune system condition,
CC especially leukaemia; for introducing exogenous nucleic acid into a
CC HSC; and for ex vivo expansion of HSCs. Also claimed are vectors,
CC host cells, and an antibody that specifically binds an HSC-specific
CC protein.
XX
SQ Sequence 644 AA;
XX
Query Match 13.3%; Score 266.5; DB 21; Length 644;
Best Local Similarity 27.0%; Pred. No. 7.7e-11;
Matches 111; Conservative 40; Mismatches 119; Indels 141; Gaps 28;
QY 12 LPELLLPAPPEAKK--PTPC-----HRCGLVDKFN-QGMVDTAKNFGGNTAME 61
Db 143 lldsltpshlpkwhespcgtpeapgnslgflckfnfkymcrlalg-gpgrvtyt 201
QY 62 ---EKTLSKSESEIRLLEILEGLC-----ESSDFECNOMLEAQBHLBAMWLQKSEY 112
Db 202 tpfqatsslea--vpfasvanvacgdeaksethylene-----kt 241
QY 113 PDLEFMPCKYKTLKVCSPGTYGPDCLACGGSGRPCS-GNGHSGSGSROGDGS--CRCH 169
Db 242 pglfhw-----gssgplcvspkfg-----csfnnggcqgdcfegdgdsfrgcgr 285
QY 170 MGYO-----GPLCTD-----CMDGYFSSLRNETHSI--- 195
Db 286 pglrllldlvtcasrnpssnptctggmchsvplsenytrcpsyql-d-sgvyhcvdld 344
QY 196 -C--TACDESKCTKSGLTNRDCECEVGVWLD--EGACVDVDECAEPPPCSAOFCNK 249
Db 345 ecqdspcaqdcvntlg--sfhc-ecwvygypspkceacedvdecaaanpc--agcgin 399
QY 250 ANGSTYCECDSSCGCTGEGPNCCKECLISGY--AREHGQCADVDECSLAE-KTCVRKN 305
Db 400 tdgsfyc-----scke-----gyivsgedstqcedidcedsdargnpc---d 437
QY 306 ENCYNTPGSVVCVCPDGFETEDACVPAEAETEG---ESPQLPSRED 352
Db 438 slcfnldgsfrgcgppgwe-----lapngvfcsrtglvsefpaprpqked 482

RESULT 46
AA78887
ID AA78887 standard; protein; 1121 AA.
XX
AC AA78887;

```

XX 14-DEC-2001 (first entry)
XX Human fibrillin 3.
DE
XX
XX Human; fibrillin; gene therapy; Marfan's syndrome.
OS Homo sapiens.
XX JP2001245664-A.
XX
XX 11-SEP-2001.
XX
XX 06-MAR-2000; 2000JP-0060009.
XX
XX 06-MAR-2000; 2000JP-0060009.
XX
XX (KAZU-) 2H KAZUSA DNA KENKYUSHO.
XX
XX WPI: 2001-610073/70.
XX N-PSDB; AA169907.
XX
XX New human fibrillin 3 gene for treating and preventing diseases showing
PT symptoms similar to Marfan's syndrome -
XX
XX Claim 1; Page 11-14; 17pp; Japanese.
XX
XX The present sequence is the protein sequence for human fibrillin 3.
CC Fibrillin 3 and its coding sequence can be used for the treatment and the
CC prevention of diseases showing symptoms similar to Marfan's syndrome.
XX
XX Sequence 1121 AA:

Query Match 13.0%; Score 260; DB 22; Length 1121;
Best Local Similarity 23.2%; Pred. No. 4e-10;
Matches 96; Conservative 32; Mismatches 120; Indels 166; Gaps 20;

QY 69 ESSEIR-----LLEFLLEGICSSDPECNQMLEAQEHEHLEAMWLQKSEFPLFEMFCVKTL 124
DB 603 echdlitgpgcfaevlqmcfs-----lassse-----avtra 634
QY 125 KVCSSPQT-YGPDLAC-----QGGSQRPSCGNGHSGDQ---SR 160
DB 635 ecccggyrgygcclcpqlpqtasayrkicphgsgytaegrdvdecmlahlcahgectins 694
QY 161 QGSGSCRHNGY-QGPLCTDCMCGYSSSLRNETHSITCTACDESKT-----CS----- 207
DB 695 lgsfrchcgggyrpdattatctld-----ndecsqvypkpcctficknktgssflscprgyl 748
QY 208 ----GLTNRCGE-----CEVGVNLDGACVDVDECAAEPPPC 241
DB 749 leedgyltckldectsrqhmcglfvcntvgaftccrppgftqhqaefndecsaagpcc 808
QY 242 SAAQFCNKANGSYCEEC-----SSCYVCTG---EGFGNCK-----ECIS 279
DB 809 gahghchmnpqstfrc-echgqfclvsqhgcednecdgphrcqhngqnlgyrscpcq 867
QY 280 GYAR---EHGCAVDVDECSLAETKCVKKNENCYTNGSYVCVCPDGF-----ETEDA 329
DB 868 gftcghsqwagcvdeneacalspctc---gsaascrltlgfrfcvcpqsfafdaqlgacqdvde 925
QY 330 CV-----PPAEAEATGEG-----SPRLPSPREDL 353
DB 926 cagrrpcpsyscantpvgflicgpcpgyfragqghcvsqglfispqpdtpkde 979

RESULT 47
AA70551
ID AA70551 standard; Protein; 1208 AA.
XX
XX AC AA70551;
XX

DT 04-JUL-2000 (first entry)
XX
XX Human latent transforming growth factor-beta binding protein 3 (1).
DE
XX
XX Human latent transforming growth factor-beta binding protein 3; hTBP-3;
KW TGF-beta inhibitor; proliferative; anti-proliferative; cytosolic;
KW cardiant; anti-inflammatory; cerebroprotective; immunosuppressive;
KW thrombolytic; osteopathic; vulnerrary; tranquilizer; antibacterial; PAI-1
KW plasminogen activator inhibitor-1; thrombomodulin; tissue growth; tumour
KW prevention; treatment; cancer; fibrosis; dilated cardiomyopathy; injury;
KW osteoporosis; myocardial infarction; congestive heart failure; sepsis;
KW thrombosis; stroke; systemic inflammatory response syndrome;
KW septic shock; sepsis syndrome; multiple organ dysfunction syndrome;
KW atherosclerotic plaque rupture.
XX
XX Homo sapiens.
XX
XX WO200012551-A1.
XX
XX 09-MAR-2000.
XX
XX 30-AUG-1999; 99WO-US19436.
XX
XX 01-SEP-1998; 98US-0098766.
XX
XX (ELIL) LILLY & CO ELI.
XX
XX Edmonds BT;
XX
XX WPI: 2000-256589/22.
XX N-PSDB; AA251928.
XX

PT Human latent transforming growth factor (TGF)-beta binding protein 3,
PT nucleic acids and vectors useful as modulators of TGF-beta, for
PT inhibiting tissue or tumor growth, and treating e.g. osteoporosis and
PT myocardial infarction -
XX
XX Claim 1; Page 61-66; 78pp; English.

XX The present sequence is a human latent transforming growth factor
CC (TGF)-beta binding protein 3 (hTBP-3) which is expressed in various
CC tissues e.g. cartilage, bone, brain, kidney, ovary, pancreas, stomach
CC and spleen. The protein is often co-expressed with TGF-beta and functions
CC as its inhibitor. hTBP-3 is useful for inhibiting or stimulating tissue
CC growth in vitro or in vivo and for inhibiting tumour growth. The hTBP-3;
CC or its homologues, and antisense nucleic acid sequences can be used to
CC regulate TGF-beta activity, especially plasminogen activator inhibitor-1
CC expression, activity or secretion, thrombomodulin expression or
CC activity, TGF-beta secretion and cellular proliferation. Modulation of
CC hTBP-3 is useful for prevention and/or treatment of diseases arising
CC from cellular effects induced by TGF-beta, especially cancer, fibrosis,
CC osteoporosis, myocardial infarction, congestive heart failure, dilated
CC cardiomyopathy, deep venous thrombosis, disseminated intravascular
CC thrombosis, stroke, sepsis, injuries involving major tissue damage and
CC trauma, systemic inflammatory response syndrome, sepsis syndrome,
CC plaque rupture.

XX Sequence 1208 AA:

Query Match 12.8%; Score 256.5; DB 21; Length 1208;
Best Local Similarity 26.8%; Pred. No. 7.6e-10;
Matches 84; Conservative 32; Mismatches 91; Indels 107; Gaps 20;

QY 128 CSPG-----TYGPPDLACQGSGSQRPSC-GNGHCSGDSRGDSCRCRHNGYQ----- 173
DB 553 cnpqyrszhpqhyrcvdenecae---epcgprgjcmtg---gsynhcnrgrylhvgag 606
QY 174 GPLCTD-----CMDGIFS-----SLRNETHSITCTACDESKTCSGL 209
DB 607 grscvdlneacaphlccdgagfcclnfpghkncycpysyrlkasrppvcddide-crdbssc 665

PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
 PT disorders (e.g. jaundice) -
 XX
 PS Disclosure; Page 236-239; 262pp; English.
 XX
 CC The present invention relates to cDNAs encoding TANGO 244,
 CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.
 CC The nucleic acids, proteins and protein modulators are useful for
 CC treating colonic disorders, inflammatory diseases, tumors,
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,
 CC allergic diseases, cardiovascular diseases, brain disorders,
 CC degenerative diseases placental, pancreatic, skeletal and muscle
 CC disorders.
 CC
 XX Sequence 1260 AA;
 SQ

Query Match 12.8%; Score 256.5; DB 22; Length 1260;
 Best Local Similarity 26.8%; Pred. No. 8e-10;
 Matches 84; Conservative 32; Mismatches 91; Indels 107; Gaps 20;

QY 128 CSPG-----TYGPDCLACGGSGRPCS-GNGHCSGDSRQDGSRCRCHMGYQ----- 173
 Db 558 cnpyrshpqrhyrcvdynecea---epcpgrgicmntyg---gsynchcnrgyrlhygag 611
 QY 174 GPLCTD-----CMNGYFS-----SLRNETHSICTRACDESCCTCGL 209
 Db 612 grscvdlnecaakphlccgagfclnfpyhkcncypgyrlkasrppvcedide-crtpssc 670
 QY 210 TNRDCE-----CEVGW-VLDEGACVDVDECAAPPCSAOFCKNANGSYTC--- 256
 Db 671 pdgkcentkpgsfkclacqpyrsgqgacrdvnece-aegspcspg-wcenlpgsftrcca 728
 QY 257 -----EECDSSCV---GCTGEGPGNCK-ECISGY--AREHGOCAVDDECS 295
 Db 729 qgyapapdgrscldvdeceagdvcdnglcsntpgsfqgcslsyhlrsdrshcedidecd 788
 QY 296 LAEKTQVRKNENCYNTPGSYVCVCPDGF-----EEEDACVPAEA----- 337
 Db 789 f-paaci--ggdcintnsgyrlcpgqnrlyggrkxqgldccsqdpslclphgacknlqg 845
 QY 338 -----ATEGESPTQ 346
 Db 846 syvcvdegfctptq 859

RESULT 50
 AAB61419
 ID AAB61419 standard; protein; 1289 AA.
 XX

AC AAB61419;

DT 04-APR-2001 (first entry)

DE Human TANGO 275 protein.

XX
 KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;
 KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;
 KW pancreatic; skeletal; muscle.
 XX

OS Homo sapiens.

XX
 PN W0200100672-A1.

XX
 PD 04-JAN-2001.

XX
 PF 29-JUN-2000; 2000MO-US18184.

XX
 PR 29-JUN-1999; 99US-0342687.

XX
 PA (MILL-) MILLENNIUM PHARM INC.

XX
 PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;
 XX
 DR WPI: 2001-050127/06.
 XX

PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
 PT disorders (e.g. jaundice) -
 XX

PS Claim 1; Fig 9; 262pp; English.

XX
 CC The present invention relates to cDNAs encoding TANGO 244,
 CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.
 CC The nucleic acids, proteins and protein modulators are useful for
 CC treating colonic disorders, inflammatory diseases, tumors,
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,
 CC allergic diseases, cardiovascular diseases, brain disorders,
 CC degenerative diseases placental, pancreatic, skeletal and muscle
 CC disorders.
 CC
 XX Sequence 1289 AA;
 SQ

Query Match 12.8%; Score 256.5; DB 22; Length 1289;
 Best Local Similarity 26.8%; Pred. No. 8.2e-10;
 Matches 84; Conservative 32; Mismatches 91; Indels 107; Gaps 20;

QY 128 CSPG-----TYGPDCLACGGSGRPCS-GNGHCSGDSRQDGSRCRCHMGYQ----- 173
 Db 587 cnpyrshpqrhyrcvdynecea---epcpgrgicmntyg---gsynchcnrgyrlhygag 640
 QY 174 GPLCTD-----CMNGYFS-----SLRNETHSICTRACDESCCTCGL 209
 Db 641 grscvdlnecaakphlccgagfclnfpyhkcncypgyrlkasrppvcedide-crtpssc 699
 QY 210 TNRDCE-----CEVGW-VLDEGACVDVDECAAPPCSAOFCKNANGSYTC--- 256
 Db 700 pdgkcentkpgsfkclacqpyrsgqgacrdvnece-aegspcspg-wcenlpgsftrcca 757
 QY 257 -----EECDSSCV---GCTGEGPGNCK-ECISGY--AREHGOCAVDDECS 295
 Db 758 qgyapapdgrscldvdeceagdvcdnglcsntpgsfqgcslsyhlrsdrshcedidecd 817
 QY 296 LAEKTQVRKNENCYNTPGSYVCVCPDGF-----EEEDACVPAEA----- 337
 Db 818 f-paaci--ggdcintnsgyrlcpgqnrlyggrkxqgldccsqdpslclphgacknlqg 874
 QY 338 -----ATEGESPTQ 346
 Db 875 syvcvdegfctptq 888

Search completed: September 10, 2002, 11:09:21
 Job time: 463 sec

